



Universidad  
Carlos III de Madrid

**Departamento de Teoría de la Señal y Comunicaciones**

**Proyecto fin de carrera**

# **Diseño y Evaluación Empírica de un Sistema de Decisión Colectiva**

*Autor: Adrián Granda Ramírez*

*Tutor: Carlos Bousoño Calzón*

**Leganés, 29 de Octubre de 2012**



# Agradecimientos

Aunque quisiera, no podría acordarme de todos los que de una manera u otra me habéis estado apoyando y ayudando para poder sacar esto adelante, pues ha sido un camino largo y difícil en el que sólo la constancia y las fuerzas que me dabais han hecho posible que llegue a su fin.

En primer lugar dar las gracias a Carlos por haberme dado la oportunidad de trabajar en su idea y poder realizar mi Proyecto Fin de Carrera, agradecerle su esfuerzo y el tiempo dedicado.

A mis padres Amando y Ana, ya que desde que dije que quería ser ingeniero me han apoyado sin condición a pesar de los buenos o malos resultados. Papá, Mamá, porque me habéis enseñado que con esfuerzo todo sale adelante, habéis confiado en mí y podré cumplir lo que por circunstancias de la vida vosotros no pudisteis.

A mi hermano Alejandro, que siempre se ha preocupado por cómo me iba y nunca le ha faltado un “suerte” antes de un examen, aunque no estoy convencido de si era porque quiere echarme rápido de casa y apoderarse de mi habitación. Si es así... ¡Alex, ya falta menos!.

A mis abuelos Augurio, Trinidad y Teresa, porque no ha habido día que me vieran y no preguntaran cómo me iba, por esas regañinas al suspender aun tratándoles de explicar que en estas carreras era algo habitual sobre todo en los primeros años. A ti, que desde donde quiera que estés, sé que me has empujado para lograrlo como el que más.

A mi familia, que a pesar de las bromas con los malos resultados, también me han felicitado por los buenos y me han apoyado en todo momento.

A todos mis compañeros de universidad, en especial a Luis, Jesús y Rocío que hemos estado juntos desde el inicio. Por todo lo que hemos sufrido juntos, lo que nos hemos ayudado no sólo académicamente, por la alegría al ver un simple 5.0 junto a nuestro nombre y sobre todo porque sois más que compañeros, os habéis convertido en mis amigos.

A mis amigos de toda la vida, porque soy consciente de que en cientos de ocasiones los he tenido abandonados y me he perdido grandes momentos con ellos debido a esto, pero a pesar de todo me han comprendido y apoyado sin dudarlo. Sé que no podréis creerlo, pero ¡por fin señores!.

A alguien a quien sin su esfuerzo estoy plenamente convencido que no podría estar escribiendo estas palabras.

A todas esas personas que sin nombrarlas, han sido muy importantes para mí, sin su cariño y apoyo este camino habría sido sin duda, mucho más difícil.

Gracias a todos por formar parte de mi vida. Pronto podré decirlo, lo he cumplido. Soy Ingeniero.



# Resumen

Basándose en la sabiduría de las multitudes, este Proyecto Fin de Carrera valora la eficacia de ciertos sistemas de decisión colectiva a través de un experimento concreto en el ámbito de la predicción de resultados quinielísticos.

Con este propósito se diseña un portal web dinámico y una interfaz de recogida de predicciones, así como diferentes procesos de agregación de información que son evaluados mediante los resultados de los encuentros de fútbol acontecidos durante el transcurso de dicho experimento.



# Abstract

Based on the wisdom of crowds, this Master Thesis Project aims to assess the efficacy of collective decision systems through a specific experiment in the field of making predictions for the football pools results.

A dynamic website and an interface for collecting predictions have been designed for this purpose, as well as different information aggregation processes that are evaluated through the final results of the football games happened during this experiment.





# Tabla de Contenidos

Índice de figuras .....	XI
-------------------------	----

Glosario .....	XIII
----------------	------

## CAPÍTULO 1

INTRODUCCIÓN .....	1
--------------------	---

1.1 Motivación del proyecto .....	1
1.2 Objetivos .....	3
1.3 Fases del desarrollo .....	3
1.4 Contenido de la memoria.....	5

## CAPÍTULO 2

ESTADO DEL ARTE DE LOS SISTEMAS DE DECISIÓN SOCIAL	7
--	---

2.1 Tipos de sistemas de decisión .....	8
2.2 Sistemas de decisiones Bayesianas.....	9
2.3 Inteligencia Colectiva.....	10
2.3.1 Introducción.....	10
2.3.2 Inteligencia Colectiva Estadística .....	11
2.3.2.1 El Colectivo .....	12
2.3.2.2 El mecanismo de agregación .....	14
2.3.2.3 Decisiones Independientes.....	15
2.3.2.4 Multitud razonablemente “inteligente” .....	16
2.3.2.5 Diversidad .....	16

2.3.2.6	Complejidad.....	17
2.3.2.7	Descentralización.....	17
2.4	Aplicación al proyecto .....	18
<b>CAPÍTULO 3</b>		
<b>TECNOLOGÍAS DE DESARROLLO DE SITIOS WEB .....</b>		<b>21</b>
3.1	Hosting y contexto general.....	21
3.2	Entorno de desarrollo web.....	23
3.3	Tecnologías.....	24
3.3.1	HTML .....	25
3.3.2	CSS.....	26
3.3.3	MySQL.....	27
3.3.4	PHP .....	28
3.4	Web Analitics.....	30
3.5	Otras herramientas.....	31
3.5.1	Editor de código fuente .....	31
3.5.2	Navegador web.....	32
3.5.3	Programa de interpretación de datos. ....	33
3.6	Selección y justificación para el desarrollo del PFC.....	33
3.7	Uso de las herramientas de desarrollo web.....	40
3.7.1	WAMP .....	40
3.7.2	Panel de control 1&1 .....	43

## **CAPÍTULO 4**

### **DISEÑO Y MEDIDAS DE CALIDAD DEL SISTEMA .....45**

- 4.1 Tipos de “sites” de quinielas: información y procesado..... 45
- 4.2 Diseño propuesto: interfaz y procesado básico ..... 46
- 4.3 Métricas de calidad: compromiso entre usabilidad e información..... 51

## **CAPÍTULO 5**

### **PROCESADO DE DATOS, EVALUACIÓN Y MODIFICACIÓN DEL SISTEMA.....53**

- 5.1 Descripción de los datos recogidos..... 53
- 5.2 Evaluación de los datos: el papel de la interfaz..... 57
- 5.3 Análisis de los datos obtenidos ..... 58
  - 5.3.1 Resultados de aciertos y participación ..... 58
  - 5.3.2 Datos de marketing y visitas web ..... 63
- 5.4 Corrección de sesgos en la recogida de información ..... 65
- 5.5 Valoración de los mecanismos de los algoritmos de decisión colectiva a la vista  
de los datos ..... 68

## **CAPÍTULO 6**

### **CONCLUSIONES Y DISCUSIÓN.....71**

## **APÉNDICE A**

### **ESQUEMA DE LA WEB.....75**

## **APÉNDICE B**

### **INFORMES DE GOOGLE ANALYTICS.....81**

### **Bibliografía .....87**



# Índice de figuras

<i>Figura 1: Ejemplo inteligencia colectiva.</i> .....	13
<i>Figura 2: Esquema de la separación de los contenidos y su presentación.</i> .....	27
<i>Figura 3: Menú WAMP.</i> .....	41
<i>Figura 4: Ejemplo de boleto de predicción.</i> .....	48
<i>Figura 5: Ejemplo de resultados de una jornada.</i> .....	50
<i>Figura 6: Triángulo de probabilidades <math>[P_1, P_X, P_2]</math>.</i> .....	52
<i>Figura 7: Probabilidades asignadas a los tres resultados según la predicción.</i> .....	54
<i>Figura 8: Constelación de puntos de decisión.</i> .....	55
<i>Figura 9: Ejemplo de probabilidades resultantes en una jornada.</i> .....	56
<i>Figura 10: Aciertos obtenidos y participación de usuarios en cada jornada.</i> .....	59
<i>Figura 11: Resumen acertantes con 10 o más aciertos.</i> .....	61
<i>Figura 12: Resumen acertantes con más de 10 aciertos.</i> .....	61
<i>Figura 13: Ranking de las 8 primeras jornadas de la temporada 12/13.</i> .....	62
<i>Figura 14: Visitantes diarios de la web durante el periodo (09/04/12 – 14/10/12).</i> .....	63
<i>Figura 15: Visitantes diarios de la web durante el periodo (14/04/12 – 15/05/12).</i> .....	64
<i>Figura 16: Visitantes diarios de la web durante el periodo (12/08/12 – 14/10/12).</i> .....	65
<i>Figura 17: Esquema de <a href="http://www.petaso1x2.com">www.petaso1x2.com</a>.</i> .....	77



# Glosario

BBDD	—	Bases de Datos.
BDD	—	Base de Datos.
CSS	—	Cascading Style Sheets.
GATC	—	Google Analytics Tracking Code.
GPL	—	General Public Licence.
GUI	—	Graphical User Interface.
HTML	—	Hypertext Markup Language.
HTTP	—	HyperText Transfer Protocol.
LAE	—	Loterías y Apuestas del Estado (España).
LFP	—	Liga de Fútbol Profesional (España).
MySQL	—	My Structured Query Language.
PC	—	Personal Computer.
PHP	—	Hypertext Transfer Protocol .
RNA	—	Redes Neuronales Artificiales.
SO	—	Sistema Operativo.
UC3M	—	Universidad Carlos III de Madrid.
W3C	—	World Wide Web Consortium.
XHTML	—	eXtensible Hypertext Markup Language.





# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1 Motivación del proyecto

La definición e importancia de la toma de decisiones. Barnard y otros teóricos posteriores tales como James March, Simon Herbert, y Henry Mintzberg, sentaron las bases para el estudio de la toma de decisiones empresariales. Pero la toma de decisiones en el seno de las organizaciones es sólo una variación de la corriente del pensamiento que fluye de vuelta a una época en la que el hombre, frente a la incertidumbre, buscaba su guía en las estrellas. Las preguntas de quién y cómo toma las decisiones, han dado forma a los sistemas mundiales de gobierno, la justicia y el orden social. "La vida es la suma de todas sus opciones," nos recuerda Albert Camus. La historia, por extrapolación, es igual a las decisiones acumuladas de toda la humanidad.

Desde los 90's, se han empezado a utilizar los sistemas predicción social como herramienta en la toma de decisiones. James Surowiecki en su libro *Wisdom and crowds* demuestra el potencial como recurso que un grupo de personas podría llegar a tener si es bien aprovechado. Él explica: "Los grupos son extraordinariamente inteligentes, y con frecuencia son más inteligentes que las personas más inteligentes de ellos".

Internet abre una vía para el empleo fácil de la inteligencia colectiva, pero su empleo implica investigar cómo hacer un uso eficiente de los diferentes recursos necesarios entre los que destacan: la gestión del colectivo humano de decisores individuales así como los mecanismos de agregación de información.

Para estudiar estos factores en un ámbito académico con escasos recursos, se recurre a un tema de predicción popular para el que se requiera un conocimiento genérico y sea capaz de suscitar la atención continuada de un colectivo de fácil acceso. Con este propósito, se escoge la predicción de quinielas... un juego basado en predecir el resultado de 15 partidos, pronosticando el equipo que ganará el encuentro o si este terminará en empate a través de una simple combinación: 1, X o 2 que representan respectivamente, victoria local, tablas y victoria visitante. Por norma general está compuesta por partidos de primera y segunda división española de la LFP, pero en determinadas ocasiones puede variar, ya siendo partidos de competiciones europeas, categorías inferiores o de selecciones. Se puede consultar la normativa aquí: [\*Normativa vigente de La Quiniela\*](#) (1).

Por todo ello, el propósito del presente proyecto es realizar un experimento que consiste en la recogida de las predicciones de los participantes durante un conjunto de jornadas de la liga española y el análisis de los mismos.

Para ello, se desarrolla una página Web en la que los usuarios participan semanalmente aportando una previsión de La Quiniela; se diseña una interfaz para cuantificar la información de forma que resulte atractiva para el usuario, la cual es procesada de forma simple para generar un único pronóstico que trata de obtener los máximos aciertos posibles.

El estudio consiguiente optimiza y valora la influencia de la interfaz en la adquisición de información y permite extraer unas primeras conclusiones de la efectividad de esta metodología de predicción para proponer recomendaciones y mejoras en su uso.

## 1.2 Objetivos

El objetivo fundamental de este Proyecto Fin de Carrera es realizar una valoración preliminar de los mecanismos de decisión colectiva empleando como caso de predicción los pronósticos de La Quiniela. Esta valoración se concluye con unas recomendaciones para el diseño de estos sistemas.

Como sub-objetivos para conseguir el principal, se proponen:

- El desarrollo de una página web para la obtención de datos.
- Una propuesta de la interfaz de recogida de información que resulte atractiva para su uso repetido.
- El desarrollo incremental de algoritmos simples de decisión colectiva. Estos algoritmos son simples agregaciones de las predicciones del colectivo humano que corrigen desviaciones de la interfaz utilizada.
- Una gestión de la información a lo largo de varias jornadas con especial énfasis en la motivación del colectivo humano.
- Una valoración, a través de los datos experimentales y sistemas de predicción alternativos, de la eficacia de la predicción colectiva y la propuesta de recomendaciones para el diseño de estos sistemas.

## 1.3 Fases del desarrollo

Las fases en el desarrollo del proyecto son:

- 1) El planteamiento inicial para que este proyecto cumpla los objetivos propuestos comenzará con un estudio y aprendizaje de las herramientas necesarias para el desarrollo de este, las cuales son indispensables para poder implementar una página web así como administrarla, gestionarla, etc.
- 2) Una vez adquiridos los conocimientos suficientes sobre las herramientas de desarrollo, nos apoyamos en ellas para diseñar un esquema de cómo será la interfaz, distribución, etc. así como los contenidos y reglas del juego que tendrán lugar en la propia web.

- 3) Concluida la fase previa de diseño comenzamos a implementar y elaborar el esquema de la web en un entorno de desarrollo local, elaborando cada una de las páginas por las que estará compuesta dicha web.
- 4) Una vez concluida una primera fase del desarrollo introducimos la parte dinámica de la web, es decir, todos los elementos de uso privado de los usuarios, como son el registro de dichos usuarios, login, introducción de pronósticos, consulta de resultados anteriores, ranking de aciertos, etc..
- 5) Cuando la versión beta de la web está concluida procedemos a alojarla en el hosting correspondiente y adaptar los pequeños errores procedentes del cambio desde un entorno local debido a la diferente interpretación del servidor web de algunos parámetros.
- 6) Una vez solucionados dichos errores procedemos a publicar nuestra página web abriéndola al público y publicitándola de diversos modos para comenzar a obtener usuarios potenciales que colaboren en la aplicación y así empezar a recabar pronósticos para la elaboración de resultados.
- 7) Tras lograr tener una interfaz adecuada para la obtención de pronósticos, aplicar el algoritmo inicial y comprobar sus resultados, trataremos de mejorar dicho algoritmo ya sea la forma de procesar los datos o la interfaz propuesta inicialmente. Una vez hecha la mejora, volvemos a analizar los resultados e intentar seguir mejorándolo hasta obtener un resultado óptimo.
- 8) Por último, una vez desarrollado por completo dicho algoritmo, procederemos a la comprobación final de resultados y afirmación o negación de si se cumplen las premisas de los teoremas utilizados.

## 1.4 Contenido de la memoria

La presente memoria comienza con una introducción de la motivación, objetivos y de la metodología empleada en la realización del proyecto. Los capítulos posteriores se organizan como sigue:

- En el capítulo 2 se expone el estado del arte del proyecto donde se comentan las diferentes teorías utilizadas para la elaboración del proyecto y la aplicación directa a este.
- En el capítulo 3 se describen los elementos, conocimientos, herramientas y tecnologías necesarias para el desarrollo de un sitio web como el necesario para este proyecto, así como la selección y justificación que hemos realizado en este caso. De manera adicional se detalla el uso de alguna de estas herramientas.
- En el capítulo 4 se evalúan las diferentes maneras en cuanto al diseño y medidas de calidad del sistema de recogida de información. Se comparan diferentes páginas web basadas en la predicción de quinielas, así como sus interfaces. De manera adicional se plantea un compromiso entre usabilidad hacia el usuario y la cantidad de información que es posible recabar.
- En el capítulo 5 se expone de manera amplia la propuesta final de recogida de pronósticos por parte de los usuarios, donde se especifica de qué manera van a ser tratados esos datos así como su razonamiento lógico. De manera adicional se enumeran los factores que creemos fundamentales para lograr una buena interfaz de cara al usuario con el propósito de conseguir la mayor cantidad de pronósticos posibles y que a su vez estos sean de calidad. Complementariamente, se hace un análisis de los sesgos observados tanto en la introducción de pronósticos por parte de los usuarios, como en la elaboración del resultado final a través de nuestro decisor.

- En el capítulo 6 se hace un extenso análisis de los resultados obtenidos en las diferentes jornadas, como la comprobación en cada caso de si ha sido posible conseguir mejorar el algoritmo a través de las innovaciones implementadas, o si se ha conseguido mejorar los resultados obtenidos por el resto de usuarios, tanto marginal como colectivamente. Adicionalmente se razona desde el punto de vista del marketing a los datos de tráfico web obtenidos y a qué patrones responden.
- En el capítulo 7 se enumeran las referencias que se han utilizado a lo largo del proyecto.

# **CAPÍTULO 2**

## **ESTADO DEL ARTE DE LOS SISTEMAS DE DECISIÓN SOCIAL**

Los seres humanos toman decisiones en condiciones de incertidumbre continuamente, es más, la incertidumbre forma parte de la vida diaria de las personas, un ejemplo tan cotidiano como el de ir tarde al trabajo y preguntarse si tomar el bus o un taxi, la persona decide con incertidumbre, porque no sabe si el bus se demorará o no, o si estará lleno, por otro lado el taxi le costará más, pero debe tomar una decisión que lo lleve a su trabajo. Este tipo de decisiones se presenta a menudo tanto en la vida cotidiana, como en el ámbito científico o en el empresarial; aunque la incertidumbre esté presente en todos esos campos, el ser humano ha desarrollado técnicas y teorías como la probabilidad y estadística para enfrentar este tipo de decisiones, pero además de estos métodos, el rápido desarrollo de dichos sistemas ha permitido fusionar todo el conocimiento en aplicaciones que permiten dar soluciones que aunque no son perfectas tienen un soporte teórico que le da mayor probabilidad de éxito, que a una decisión tomada al azar (2).

Por todo ello, y para poder hallar una solución coherente en la decisión de un algoritmo que recoja la información obtenida y la aplique intentando maximizar las probabilidades de éxito, nos basaremos en los que hemos considerado que serán los pilares de nuestro decisor como posteriormente expondremos: la *Decisión Bayesiana* y la *Inteligencia Colectiva*.

## 2.1 Tipos de sistemas de decisión

En la actualidad, hay infinidad de sistemas de decisión que se usan para intentar prever un resultado, ya sea deportivo, cotización de acciones, etc. Estos sistemas abarcan una amplia variedad de decisores, ya que se basan en muy distintos parámetros para intentar optimizar el resultado final. A continuación expondremos alguno de los ejemplos más significativos en el ámbito de los sistemas de decisión:

- **Decisión por consenso:** se denomina así al proceso de decisión que busca no solamente el acuerdo de la mayoría de los participantes, sino también persigue el objetivo de resolver o atenuar las objeciones de la minoría para alcanzar la decisión más satisfactoria.
  
- **Decisión por procesos estructurados:** se denomina así a los sistemas que basan sus decisiones en eventos históricos para argumentar su pronóstico. Un claro ejemplo de ello y muy en la línea de nuestro proyecto, fue el estudio realizado en la Universidad de Guadalajara, México, (3) donde se utilizaron redes neuronales artificiales (RNA), en concreto el perceptrón multicapa, que analizaba para cada partido datos tales como: presupuesto de los equipos, clasificación actual, puntos conseguidos en los últimos cuatro partidos, goles anotados y encajados en los últimos cuatro partidos, etc. El resultado del experimento fue que pronosticando resultados de los partidos de fútbol de primera división de la LFP durante cuatro jornadas. Sus resultados finales con su mejor aproximación fueron el 53'75% de los pronósticos de media acertados.



- **Decisión por procesos cognitivos (4):** este tipo de sistema de decisión es el utilizado normalmente por las empresas a la hora de tomar una decisión. Está basada en los siguientes procesos: Observación, comparación, codificación, organización, clasificación, resolución, evaluación y retroalimentación.
- **Decisión por inferencia bayesiana (5):** es un tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis pueda ser cierta, es decir, usa un estimador numérico del grado de creencia en una hipótesis aún antes de observar la evidencia y calcula un estimador numérico del grado de creencia en la hipótesis después de haber observado la evidencia.

## 2.2 Sistemas de decisiones Bayesianas (6)

Considerando la posibilidad de un evento incierto, por ejemplo, que la capa de hielo del Ártico habrá desaparecido a finales de siglo, no es un evento que se pueda repetir varias veces a fin de definir una noción de probabilidad como si se podría hacer por ejemplo al tirar un dado y observar que cara queda visible.

Sin embargo, por lo general se tiene alguna idea, por ejemplo, de la rapidez con la que pensamos que el hielo polar se está derritiendo. Si ahora obtenemos nuevos elementos de prueba, por ejemplo, una nueva forma de observación satelital de la Tierra con recopilación de información de diagnóstico, podremos revisar nuestra opinión sobre la tasa de pérdida de hielo. Nuestra evaluación de tales asuntos afectará a las acciones que tomamos, como por ejemplo, el grado en que nos esforzamos para reducir la emisión de gases de efecto invernadero. En estas circunstancias nos gustaría ser capaces de cuantificar nuestra expresión de la incertidumbre y hacer revisiones precisas de dicha incertidumbre a la luz de nuevos datos, así como posteriormente poder tomar decisiones óptimas en consecuencia. Todo esto se puede lograr a través de la *interpretación Bayesiana de la probabilidad*.

El uso de la probabilidad para representar la incertidumbre no es una solución específica para esto, pero es inevitable si queremos respetar el sentido común al hacer inferencias racionales coherentes. Por ejemplo, Cox (7) comprobó que si los valores numéricos se usan para representar grados de creencia, un simple conjunto de axiomas que codifiquen propiedades de sentido común de dichas creencias, nos lleva a un conjunto de reglas para manipular grados de creencia que son equivalentes a las *reglas de probabilidad de suma y producto*. Esto proporcionó la primera prueba rigurosa de la teoría de que la teoría de la probabilidad puede considerarse como una extensión de la lógica booleana para situaciones de incertidumbre (8).

## 2.3 Inteligencia Colectiva (9)

### 2.3.1 Introducción

La inteligencia colectiva es el resultado de la adecuada agregación de la información local de muchas personas para generar una óptima solución global a un problema. A menudo, estas soluciones llegan a ser mejores que las que cualquier individuo podría proporcionar (10). Los mecanismos por los que las soluciones inteligentes se generan colectivamente se denominan *sistemas inteligentes*. Todos los sistemas de inteligencia colectiva cuentan con una población de participantes (es decir, un colectivo) y un medio de la agregación de sus conocimientos en una decisión colectiva (es decir, un mecanismo de agregación). Por ejemplo, agregando individuos en un debate a través de la conversación o agregando individuos en la democracia a través del voto. En las condiciones más prometedoras, al agregar los muchos puntos de vista, los efectos de los errores individuales se reducen al mínimo y la solución colectiva derivada es relativamente óptima. En otras palabras, en la solución colectiva a problemas, ningún individuo conoce la solución, a pesar de que el grupo como un todo si pueda llegar a conocerla. Este fenómeno se conoce como inteligencia colectiva estadística y puede considerarse en su sentido coloquial como la "ley de los grandes números" (*Law of large numbers*).

### 2.3.2 Inteligencia Colectiva Estadística

Un método para generar conocimiento a partir de grupos de personas que evite el escollo de la dinámica de grupo es el fenómeno que se conoce como inteligencia colectiva estadística. Simplemente, se trata de la generación de conocimiento a través de una media ponderada de las estimaciones independientes e individuales a preguntas específicas. James Surowiecki en su libro *Wisdom and crowds* demuestra el potencial como recurso que un grupo de personas podría llegar a tener si es bien aprovechado. Él explica: "Los grupos son extraordinariamente inteligentes, y con frecuencia son más inteligentes que las personas más inteligentes de ellos" (11). Sin embargo, no todos los grupos son buenos generadores de conocimiento.

Por otro lado, los acuerdos de las multitudes son poco eficientes y peligrosos para transmitir el conocimiento de todos sus miembros. Incluso los equipos pequeños por lo general no utilizan todos los conocimientos de sus miembros debido a las dinámicas de grupo. La identidad social juega un papel importante en las interacciones grupales.

Si bien diversas perspectivas son un atributo importante de grupos efectivos, los individuos pueden sentirse demasiado diferentes y por ello modificar su comportamiento con el fin de semejarse al grupo (12). Las normas sociales presionan a los individuos para que se comporten como se espera (13). Además, las presiones en pequeños grupos para llegar a un consenso puede llevar a los individuos a buscar la conformidad a través de respuestas exactas (14).

Gustave Le Bon, el sociólogo del siglo XIX y cínico con las decisiones del grupo, llamó a estos fenómenos de "contagio" que tienen el efecto de alterar las "emociones o pensamientos particulares como resultado de ser sumergido en una multitud" (13). La dinámica de grupos grandes tiene una tendencia a inhibir la expresión de su gama completa de conocimiento.

Por lo tanto, con el fin de alcanzar su potencial para la generación de conocimiento y la resolución de problemas, un grupo debe de ser pensado como una herramienta que ha de utilizarse correctamente.

A continuación definiremos cómo funciona un colectivo a la hora de decidir sobre un determinado problema así como expondremos los diferentes criterios que debe de cumplir.

### 2.3.2.1 El Colectivo

El relato típico de resolución de problemas implica que un experto aplica sus conocimientos para generar una solución. A través de la inteligencia colectiva, sin embargo, es el mismo colectivo el que se considera el experto. El colectivo puede ser pensado como un meta-individuo que posee, genera, y actúa sobre el conocimiento de la misma manera que un ser humano.

Los colectivos son dinámicos, no tienen continuidad en la identidad a pesar de los cambios en la composición, lo que nos permite pensar en ellos como individuos persistentes. La inteligencia colectiva utiliza el conocimiento en manos de un grupo de individuos que actúan sobre el meta-nivel. El conocimiento basado en el grupo puede ser considerado un experto.

Como un experto, un colectivo tiene más conocimientos que otros individuos. Consideramos el siguiente modelo donde la diversidad en el conocimiento local genera nuevo conocimiento.

Supongamos que en una clase de 9 alumnos, han de decidir si eligen a Will Jay como delegado de clase. Los aspectos en los que se valora de manera positiva o negativa a un posible delegado son 11, representados en la tabla de la *Figura 1* por las columnas donde un 1 representa que el aspecto es positivo y un 0 que es negativo. Cada uno de los 9 alumnos sólo conoce, y por tanto valora, 3 de los 9 aspectos de Will Jay, y elige si le vota positiva o negativamente en función de la mayoría, es decir, si tiene 2 o 3 aspectos positivos le votará a favor, y si tiene 1 o ninguno, votará en contra.

	0	0	0	1	1	0	1	0	1	1	1	Outcome
Adam	0	0	0									0
Bob		0	0	1								0
Cam			0	1	1							1
Dan				1	1	0						1
Evan					1	0	1					1
Fred						0	1	0				0
George							1	0	1			1
Harry								0	1	1		1
Ivan									1	1	1	1
DECISION												1

*Figura 1: Ejemplo inteligencia colectiva.*

Como se puede observar en la primera fila de la tabla, Jay tiene 6 aspectos positivos y 5 negativos, por lo que la elección correcta es que sí debería salir elegido. Pues bien, observando a Adam, éste conoce 3 aspectos negativos, y por tanto vota en contra. Del mismo modo, Bob conoce sólo 1 aspecto positivo de Jay, por lo que también vota en contra. Por lo que si toda la clase sólo conociera estos 4 primeros aspectos de Jay, se tomaría la decisión equivocada pensando que sería un mal delegado.

Al observar el conjunto de la tabla y debido a que es un conjunto heterogéneo donde aglutinan el total de los conocimientos sobre Jay, 6 alumnos votan a favor y 3 en contra por lo que la elección es correcta y Jay sale elegido delegado de la clase, por lo que se comprueba que la suma de los conocimientos locales genera nuevo conocimiento (15). Esto demuestra que las diversas opiniones, incluso inexactas, pueden contribuir a la generación de inteligencia colectiva estadística.

La inteligencia se genera porque la suma de los diferentes conocimientos locales cubre el conjunto completo de información necesaria para resolver un problema.

Los seres humanos son increíbles sintetizadores de información del medio ambiente. La fuerza evolutiva está detrás del poder de nuestros procesos cognitivos, tenemos una capacidad muy desarrollada para el razonamiento cognitivo, el análisis y resolución de problemas.

Los seres humanos destacan, en comparación con las computadoras, en la solución de problemas mal definidos, es decir, aquellos que tienen objetivos complejos, soluciones múltiples, o una naturaleza cambiante. Estos problemas complejos requieren la aplicación de conocimientos, la intuición, el diagnóstico y el análisis. Estas habilidades se han desarrollado para que podamos predecir con exactitud y prepararnos para futuros eventos.

#### 2.3.2.2 El mecanismo de agregación

Un colectivo sin un *agregador* no es más poderoso que los individuos que lo componen.

Un *mecanismo de agregación* sirve para dos propósitos en la obtención de la inteligencia colectiva. Uno, representar la información pertinente de cada individuo en la colectividad. Dos, combinar la información de tal manera que ésta sea útil.

Los *sistemas de recomendación* hacen uso del comportamiento del usuario, ya sea implícita o explícitamente generado, como un medio para recomendar artefactos potencialmente interesantes para los usuarios en el sistema. Un algoritmo para tal recomendación es el algoritmo de filtrado colaborativo que hizo popular *Amazon.com*.

El *algoritmo de filtrado cooperativo* compara las conductas de los miembros del colectivo tales que los individuos similares se recomiendan aquellos artefactos a los que aún no han tenido acceso (16). Los sistemas de votación son utilizados por las democracias para determinar la opinión colectiva. Mientras que muchos algoritmos de agregación se han explorado en profundidad (17), la democracia directa, practicada por la antigua Atenas, es quizás la más conocida.

### 2.3.2.3 Decisiones Independientes

Los seres humanos están limitados en su capacidad de procesar toda la información disponible. Es lógico pensar que si una persona es buena en la solución de un problema particular con su información limitada, un grupo sería aún mejor. Sin embargo, esto es cierto sólo bajo ciertas condiciones. A menudo, cuando se trabaja en un grupo, las soluciones no van a mejorar. Esto es particularmente evidente en las multitudes, donde la violencia es común. En las multitudes, la gente se siente parte de un grupo y ya no responsables de sus acciones. Estos sentimientos son a menudo responsables de los disturbios y de los crímenes que cometen las bandas. Sin embargo, el fenómeno se mantiene incluso en las situaciones más tranquilas donde el pensamiento grupal puede llevar a un equipo lejos de una buena respuesta. El *pensamiento grupal* se refiere al proceso mediante el cual cada miembro del grupo cambia su opinión con respecto a la percepción de un consenso del grupo (18).

De esta manera, un buen método para obtener las predicciones de la multitud, debe ofrecer la oportunidad de evitar estas condiciones interactivas que conducen a grupo al naufragio y así aumentar al máximo el poder que tienen grupos de personas como sintetizadores de información. Un componente esencial de esta maximización es que los participantes mantienen su individualidad al tomar decisiones independientes. Los participantes deben tener la libertad de expresar sus creencias, sin sentir la influencia de los demás. Logramos esto mediante el fomento de la competencia entre los participantes, y no por consenso. Debido a la competencia, es poco probable que los participantes compartan su información privada y por lo tanto influyan en otros o se sientan presionados socialmente para modificar sus decisiones.

Los participantes en una predicción deben de ser auto-seleccionados. En otras palabras, la población se selecciona a través de decisiones de los individuos a participar. Este método de organización es, en este nivel, no competitivo, y por ello la población no sólo incluirá buenos predictores, si no a un público genérico.

#### 2.3.2.4 Multitud razonablemente “inteligente”

La inteligencia colectiva se basa en la creencia de que las personas no toman siempre decisiones perfectas.

No somos seres completamente racionales. Los seres humanos están limitados por lo que el científico de la computación y el filósofo Herbert Simon llama *racionalidad limitada*. Tal como él lo dijo, "la gente tiene límites de experiencia en la formulación y resolución de problemas complejos y de procesamiento (recepción, almacenamiento, recuperación, transmisión) de información" (19). Un ser humano es un buen, pero no ideal, solucionador de problemas. La inteligencia colectiva utiliza una mejor solución, es decir, la unidad de los participantes. Por lo tanto, un conjunto de personas tiene más conocimiento que cualquier persona sola, incluso el más experto.

Una manera con la que solemos resolver problemas complejos, es convocar a un grupo de personas con experiencia y conocimientos de un tema y permitirles desarrollar una solución conjunta. Sin embargo, para lograr éxito depende mucho de encontrar a las personas adecuadas para componer esta comisión así como crear un entorno en el que sientan que pueden expresar lo que realmente creen.

#### 2.3.2.5 Diversidad

La *diversidad* es el mecanismo fundamental que está detrás de la aparición de la inteligencia colectiva. La diversidad "constituye la base para una explicación de por qué el esfuerzo colectivo de un grupo a menudo puede superar al de un individuo: por el hecho de ser diferentes, las personas pueden mejorar mutuamente sus diferentes soluciones a un problema" (20). Mediante la utilización de una población que represente diferentes piezas de información, emerge una idea más clara de la visión completa de la solución, que es necesaria para predicciones exactas.

Con todas las maneras con las que las personas pueden diferir en una opinión, parece como si la gente rara vez expresara creencias similares entre todos. Sin embargo, es expresamente por la influencia de otros, que pueden llegar a convencerse de una opinión por consenso. Por esta razón, es necesario tener tanto una población diversa originalmente, como que se mantenga independiente de la influencia de otros.



#### **2.3.2.6 Complejidad**

Anécdotas de inteligencia colectiva sugieren que grandes grupos de personas que trabajan individualmente contienen una sabiduría que no se encuentra en una persona individual y que esta sabiduría puede ser aplicada y, de hecho sólo funciona bien, en problemas difíciles, tales como las predicciones. El objetivo es poner de manifiesto a través de la inteligencia colectiva lo que no es evidente para los individuos. Estamos particularmente interesados en cuestiones de predicción, como son los problemas complejos porque dependen de una constelación de factores, ya que es ahí donde la inteligencia colectiva puede llegar a asumir mejor todos y cada uno de dichos factores.

#### **2.3.2.7 Descentralización**

Los individuos, equipos y organizaciones no son la única manera de resolver los problemas. La inteligencia distribuida en un sistema descentralizado es una manera importante de resolver problemas y aumentar nuestro conocimiento ya que produce respuestas a las preguntas que son demasiado complejas para que un individuo o grupo pueda comprenderlas. La descentralización se refiere a la propiedad de un sistema en el que las decisiones "son hechas por individuos basándose en su propio conocimiento local y específico en lugar de por un planificador omnisciente o previsor" (11). Por tanto, lo ideal es que los individuos de la población sean libres de elegir cuándo y cómo participar.

## 2.4 Aplicación al proyecto

La inteligencia colectiva como generadora de pronósticos podría servir como un gran recurso para los individuos, las organizaciones y la sociedad. Existe información implícita simplemente a través de la correcta agregación de opiniones de los individuos. Sin embargo, el desarrollo de la confianza en la inteligencia colectiva es la principal preocupación para el éxito de esta herramienta de predicción.

Debido a que consideramos que el caso particular de La Quiniela de la LFP es un claro ejemplo en el que poder aplicar la inteligencia colectiva, intentaremos demostrar con resultados dichas teorías procurando lograr un colectivo lo más idóneo posible dentro de nuestras posibilidades.

Para poder tomar decisiones independientes, los participantes de nuestra web de predicción son auto-seleccionados. En otras palabras, la población se selecciona a través de decisiones de los individuos a participar. Este método de organización es, en este nivel, no competitivo, la población no sólo incluirá buenos predictores, si no a un público genérico.

La gente está con toda seguridad demasiado influenciada tanto social como anímicamente por los distintos medios de comunicación al pensar en una predicción futbolística, lo que hace que tiendan a hacer pronósticos basados en otros muchos factores que alteran el resultado de la multitud. Por tanto, nuestro proyecto requiere que los individuos tomen sus propias decisiones para evitar los errores de juicio que se producen a través de las pobres técnicas de concentración que se encuentran en las multitudes.

Desgraciadamente, para poder obtener una muestra lo suficientemente grande como para que nuestro sistema pueda funcionar de una manera más o menos eficiente, no podemos hacer distinciones entre los más y menos expertos en la materia, representado más o menos fielmente por los resultados obtenidos a través de las jornadas. De ser esto posible, sólo habrían de participar los que obteniendo buenos resultados están muy seguros de sus predicciones basándolas en algo más que el mero

hecho de adivinar un resultado y así se podría asegurar una multitud razonablemente “inteligente”.

En nuestro caso particular necesitamos que la población que manejemos sea lo más diversa posible en cuanto a gustos futbolísticos se refiere, es decir, si nuestros individuos se basaran básicamente en aficionados del Real Madrid, que viven en Madrid, y debido a la gran rivalidad existente, tengan un odio férreo tanto al Barcelona como al Atlético de Madrid, nuestra predicción siempre tendería hacia un resultado más positivo hacia el Real Madrid y más negativo hacia el Barcelona y Atlético de Madrid de lo que en principio se conseguiría con una muestra más homogénea. Por tanto lo ideal sería conseguir una multitud diversa en lo que se refiere tanto a gustos futbolísticos como a la geografía española ya que de esta manera se podría eliminar mejor el sesgo de una participación incondicional hacia algunos determinados equipos.

Una razón de más en la que apoyarnos para defender esta teoría, es en que anécdotas de inteligencia colectiva sugieren que ésta funciona mejor cuanto más complejo sea el problema a resolver, ya que es aquí donde un individuo tiene más problemas para analizar toda la información que percibe. Por eso este tipo de inteligencia está más orientada a problemas complejos que dependen de una constelación de factores, como en este caso los resultados futbolísticos en una determinada jornada, en la que los equipos no sólo depende de cómo jueguen el partido, si no de cómo llegan cada uno de ellos, tanto el grupo como individualmente sus jugadores, etc. y es aquí donde la inteligencia colectiva puede destacar ante el conocimiento individual.

Una vez comprendido que nuestro proyecto, es un claro ejemplo en el que la inteligencia colectiva puede dar buenos resultados basándonos en métodos como la *inteligencia colectiva estadística* donde la generación de conocimiento se efectúa a través de la media ponderada de las estimaciones independientes e individuales de la respuesta a preguntas específicas hechas a un colectivo, observamos que podríamos complementar esta teoría con un *sistema de decisiones Bayesiano*, de tal forma que asignando valores numéricos a diferentes grados de creencia sobre un resultado concreto y haciendo

que estos respondan a pensamientos del sentido común, nos llevará a un conjunto de reglas equivalente a las *reglas de probabilidad de suma y producto* (7).

Por todo ello, en nuestro caso en particular, hemos de tratar la información introducida por los usuarios de manera cuantitativa para expresar los diferentes grados en los que creen que un resultado va a producirse, de tal manera que representamos cada uno de los tres posibles resultados (victoria local, empate y victoria visitante), con tres grados diferentes de confianza en ellos y uno de completa incertidumbre en el resultado, como explicaremos en el capítulo 5, y así poder tratar estos datos de manera probabilística y obtener una predicción lo más acertada posible.

# CAPÍTULO 3

## TECNOLOGÍAS DE DESARROLLO DE SITIOS WEB

### 3.1 Hosting y contexto general

Web Hosting o alojamiento web, es el servicio que provee a los usuarios de Internet un sistema para poder almacenar información, imágenes, vídeo, o cualquier contenido accesible vía web. Es una analogía de "hospedaje o alojamiento en hoteles o habitaciones" donde uno ocupa un lugar específico, en este caso la analogía alojamiento web o alojamiento de páginas web, se refiere al lugar que ocupa una página web, sitio web, sistema, correo electrónico, archivos etc. en internet o más específicamente en un servidor que por lo general hospeda varias aplicaciones o páginas web.

Las compañías que proporcionan espacio de un servidor a sus clientes se suelen denominar con el término en inglés “web host”.

El alojamiento web aunque no es necesariamente un servicio como tal, se ha convertido en un lucrativo negocio para las compañías de internet alrededor del mundo debido a que en la sociedad actual, prácticamente cualquier tipo de empresa (incluso las más pequeñas y emprendedores individuales) tiene su *web site*, ya que es la mejor manera de poder promocionarse, darse a conocer y mostrar sus productos y/o servicios a un bajo coste.

Hosting se puede definir como el lugar donde se encuentra una página web, aunque esta definición simplifica de manera conceptual el hecho de que el alojamiento web es en realidad espacio en Internet para prácticamente cualquier tipo de información, sea archivos, sistemas, correos electrónicos, videos, etc..

El servicio mínimo que ofrecen un Hosting se basa simplemente en el almacenamiento de los archivos necesarios para implementar una página web, como son los HTML, PHP, CSS, etc. como veremos más adelante, así como las imágenes si las hubiere. Normalmente se ofrece a través de una interfaz de subida y bajada de archivos con el que almacenarlos fácilmente en el servidor del proveedor y que da a estos salida a internet. El tamaño de almacenamiento varía según el servicio contratado, tratándose normalmente desde 1GB a ilimitado.

Por norma general estos servicios tienen asociado el coste de un dominio. Un dominio es una red de identificación asociada a un grupo de dispositivos o equipos conectados a la red Internet, o simplemente, el nombre visible de una página web a ojos del usuario final. El dominio es necesario para cualquier tipo de página web, ya que es la puerta de entrada a dicha web, por tanto, si el servicio de hosting no lo incluye, es el usuario quien tiene que registrar un dominio mediante un registrador o bien usar un subdominio de la misma compañía. Un dominio web se compone por el nombre de la página seguido, separado por un punto, por el nombre de dominio de nivel superior, siendo este el que define de qué país es la página (por ejemplo 'es', 'it' y 'fr' para España, Italia y Francia respectivamente), o bien otros conjuntos como son por ejemplo 'net' para estructuras de red, 'info' para información o 'com' el genérico y más usado en la red. Un ejemplo sería *ejemplodominio.com*. Así mismo, un

subdominio sería añadir un nombre anterior al dominio, es decir, subdominio.ejemplodominio.com.

Adicionalmente como otro de los servicios primordiales, las compañías de Web Hosting suelen proveer tanto de una o varias BBDD a sus clientes donde poder almacenar toda la información que necesita la web del cliente para sus diversas funciones, como son el registro y datos de usuarios, estadísticas de uso, entradas de texto, etc. como de una interfaz donde poder almacenar, visualizar, organizar, etc. las distintas BBDD y tablas que las componen.

Dependiendo del servicio contratado, los Web Hosting suelen proporcionar cuentas de correo asociadas con el dominio contratado, de tal forma que puedas tener un correo personal de la web, por ejemplo, ejemplo@dominio.com.

### **3.2 Entorno de desarrollo web**

Para poder empezar a desarrollar una página web no es necesario utilizar un hosting directamente, si no que se puede crear un entorno web en local que ayuda a simplificar y agilizar el constante cambio de la estructura web, contenidos, diseño, etc. ya que no se ha de estar subiendo estos contenidos al servidor.

Esto se consigue a través de un entorno de desarrollo web que aúne tres elementos necesarios para ello: Intérprete de PHP, servidor local de HTTP y un sistema de gestión de BBDD.

PHP como se explicará más adelante es un lenguaje de programación interpretado para HTML diseñado para la creación de páginas web dinámicas. Dicho lenguaje se usa para la interpretación del lado del servidor, por lo que un PC normal no puede ejecutarlo. Por ello es necesario instalar un intérprete de PHP que configure un PC de manera que actúe como un servidor local y pueda interpretar dicho lenguaje.

Un servidor HTTP es un programa informático que procesa una aplicación del lado del servidor realizando las conexiones pertinentes con el cliente generando o cediendo una respuesta en cualquier lenguaje o aplicación del lado del cliente, el código recibido por el cliente suele ser compilado y ejecutado por un navegador web. Para

poder simular esto en un mismo PC, se puede instalar un Servidor Web Local que es aquel Servidor Web que reside en una red local al equipo de referencia. Los archivos se almacenan en un directorio determinado por la configuración, y es directamente modificable, lo que agiliza el continuo cambio en el desarrollo de, en nuestro caso, páginas web. Hoy en día el servidor HTTP más utilizado es el Apache, ya que se trata de un servidor de código abierto que se encuentra disponible para los tres SO más importantes, LINUX, Windows y Macintosh y por tanto será con el que más facilidades tengamos para trabajar.

Para poder tener un completo entorno de desarrollo web se necesita un sistema de gestión de BBDD local con el que poder manejar las diferentes BBDD y sus tablas dónde se almacena toda la información de la página web, como son, los registros de los usuarios y sus datos, información dinámica de la web, etc. MySQL ofrece todo esto gracias a su software libre de una manera ágil y totalmente integrado con los sistemas anteriores.

### **3.3 Tecnologías**

La solución más cercana a la hora de implementar la página web de nuestro proyecto, trata de varios lenguajes de programación tales como HTML, PHP y MySQL, debido a que se trata de herramientas de software libres que han logrado un gran desarrollo a lo largo de los años y se usan prácticamente en la totalidad de las páginas web. Con ellos se consigue diseñar la interfaz, manejar la parte interactiva de la web y gestionar las BBDD asociadas a esta de manera sencilla con unos conocimientos básicos.

Adicionalmente, existen decenas de lenguajes que mejoran y ayudan de manera considerable tanto a la simplificación como a la mejora de la web. Un claro ejemplo de ello lo encontramos en el CSS que recoge toda la parte de diseño (tamaños, colocaciones, colores) de un documento HTML, por lo que quitan mucha complejidad a la hora de diseñarlo ya que es redundante y no hay que aplicar en cada objeto todos los atributos deseados.



A continuación explicaremos de manera breve cada una de estas tecnologías aplicadas:

### 3.3.1 HTML

Definiéndolo de forma sencilla, "HTML es lo que se utiliza para crear todas las páginas web de Internet". Más concretamente, HTML es el lenguaje con el que se escriben la mayoría de páginas web. (21)

Los diseñadores utilizan el lenguaje HTML para crear sus páginas web, los programas que utilizan los diseñadores generan páginas escritas en HTML y los navegadores que utilizamos los usuarios muestran las páginas web después de leer su contenido HTML.

Aunque HTML es un lenguaje que utilizan los ordenadores y los programas de diseño, es muy fácil de aprender y escribir por parte de las personas. En realidad, HTML son las siglas de *HyperText Markup Language* (lenguaje marcado de hipertexto) o lo que es lo mismo, un lenguaje de marcado predominante para la elaboración de páginas web que se utiliza para describir y traducir la estructura y la información en forma de texto, así como para complementar el texto con objetos tales como imágenes. El HTML se escribe en forma de *etiquetas*, rodeadas por corchetes angulares (<,>). HTML también puede describir, hasta un cierto punto, la apariencia de un documento, y puede incluir un script (por ejemplo JavaScript), el cual puede afectar el comportamiento de navegadores web y otros procesadores de HTML.

El lenguaje HTML es un estándar reconocido en todo el mundo y cuyas normas define un organismo sin ánimo de lucro llamado World Wide Web Consortium, más conocido como W3C. Como se trata de un estándar reconocido por todas las empresas relacionadas con el mundo de Internet, una misma página HTML se visualiza de forma muy similar en cualquier navegador de cualquier sistema operativo. No obstante no todos interpretan de manera exacta el código generado, por lo que en la mayoría de ocasiones es necesario retocar éste para que se optimice el resultado en el mayor número de navegadores posibles, o al menos en los más usados.

El propio W3C elabora las normas que deben seguir los diseñadores de páginas web para crear las páginas HTML. Las normas oficiales están escritas en inglés y se pueden consultar de forma gratuita en las siguientes direcciones:

[Especificación oficial de HTML 4.01](#)

[Especificación oficial de XHTML 1.0](#)

[Especificación oficial de HTML 5](#)

El estándar XHTML (eXtensible HTML) 1.0 incluye el 95% del estándar HTML 4.01, ya que sólo añade pequeñas mejoras y modificaciones menores. Afortunadamente, no es necesario leer las especificaciones y recomendaciones oficiales de HTML para aprender a diseñar páginas con HTML o XHTML. Las normas oficiales están escritas con un lenguaje bastante formal y algunas secciones son difíciles de comprender. Por ello, en los próximos capítulos se explica de forma sencilla y con decenas de ejemplos la especificación oficial de XHTML.

El estándar HTML 5 todavía se encuentra en modo experimental, lo cual indica la misma W3C; aunque ya es usado por múltiples desarrolladores web por sus avances, mejoras y ventajas.

### **3.3.2 CSS**

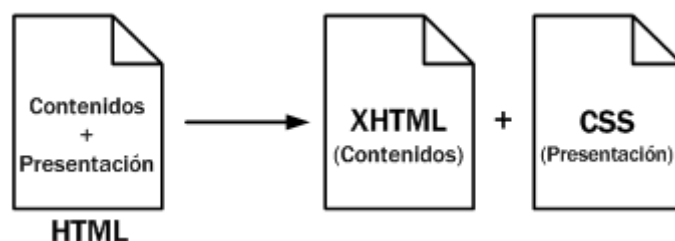
Originalmente, las páginas HTML sólo incluían información sobre sus contenidos de texto e imágenes. Con el desarrollo del estándar HTML, las páginas empezaron a incluir también información sobre el aspecto de sus contenidos: tipos de letra, colores y márgenes (21).

La posterior aparición de tecnologías como JavaScript, provocaron que las páginas HTML también incluyeran el código de las aplicaciones (llamadas *scripts*) que se utilizan para crear páginas web dinámicas.

Incluir en una misma página HTML los contenidos, el diseño y la programación complica en exceso su mantenimiento. Normalmente, los contenidos y el diseño de la

página web son responsabilidad de diferentes personas, por lo que es conveniente separarlos.

CSS es el mecanismo que permite separar los contenidos definidos mediante XHTML y el aspecto que deben presentar esos contenidos lo representamos en la *Figura 2*:



*Figura 2: Esquema de la separación de los contenidos y su presentación.*

Una ventaja añadida de la separación de los contenidos y su presentación es que los documentos XHTML creados son más flexibles, ya que se adaptan mejor a las diferentes plataformas: pantallas de ordenador, pantallas de dispositivos móviles, impresoras y dispositivos utilizados por personas discapacitadas.

De esta forma, utilizando exclusivamente XHTML se crean páginas web "*feas*" pero correctas. Aplicando CSS, se pueden crear páginas "*bonitas*" a partir de las páginas XHTML correctas.

### 3.3.3 MySQL

El sistema de base de datos operacional MySQL es hoy en día uno de los más importantes en lo que hace al diseño y programación de base de datos de tipo relacional. Cuenta con millones de aplicaciones y aparece en el mundo informático como una de las más utilizadas por usuarios del medio. El programa MySQL se usa como servidor a través del cual pueden conectarse múltiples usuarios y utilizarlo al mismo tiempo (22).

La historia del MySQL, cuya siglas significan My Structured Query Language (Lenguaje de Consulta Estructurado), se remite a principios de la década de 1980. Programadores de IBM lo desarrollaron para contar con un código de programación que permitiera generar múltiples y extendidas bases de datos para empresas y organizaciones de diferente tipo. Desde esta época numerosas versiones han surgido y muchas de ellas fueron de gran importancia. Hoy en día MySQL es desarrollado por la empresa Sun Microsystems.

Una de las características más interesantes de MySQL es que permite recurrir a bases de datos multiusuario a través de la web y en diferentes lenguajes de programación que se adaptan a diferentes necesidades y requerimientos. Por otro lado, MySQL es conocida por desarrollar alta velocidad en la búsqueda de datos e información, a diferencia de sistemas anteriores. Las plataformas que utiliza son de variado tipo y entre ellas podemos mencionar LAMP, MAMP, SAMP, BAMP y WAMP (aplicables a Mac, Windows, Linux, Perl y Python entre otras).

Se están estudiando y desarrollando nuevas versiones de MySQL que buscan presentar mejoras y avances para permitir un mejor desempeño en toda aquella actividad que requiera el uso de bases de datos relacionales. Entre estas mejoras podemos mencionar un nuevo dispositivo de depósito y almacenamiento, backup para todos los tipos de almacenamientos, replicación segura, planificación de eventos y demás.

### **3.3.4 PHP**

PHP es un lenguaje de programación interpretado para HTML diseñado originalmente para la creación de páginas web dinámicas. Se usa principalmente para la interpretación del lado del servidor (server-side scripting) pero actualmente puede ser utilizado desde una interfaz de línea de comandos (23).

PHP es un acrónimo recursivo que significa PHP Hypertext Preprocessor (inicialmente PHP Tools, o, Personal Home Page Tools). Fue creado originalmente por Rasmus Lerdorf en 1994; sin embargo la implementación principal de PHP es producida ahora por The PHP Group y sirve como el estándar de facto para

PHP al no haber una especificación formal. Publicado bajo la PHP License, la Free Software Foundation considera esta licencia como software libre.

Una vez introducido el término, ¿qué significa realmente? un ejemplo nos aclarará las cosas:

```
<html>
  <head>
    <title>Example</title>
  </head>
  <body>
    <?php
      echo "Hola, isoy un script PHP!";
    ?>
  </body>
</html>
```

En lugar de usar muchos comandos para mostrar HTML (como en C o Perl), las páginas PHP contienen HTML con código incluido en el mismo que hace "algo" (en este caso, mostrar "Hola ¡soy un script PHP!"). El código PHP está entre medio de etiquetas de comienzo y final especiales `<?php` y `?>` que nos permitirán entrar y salir del "modo PHP".

Lo que distingue a PHP de algo lado-cliente como Javascript, es que el código es ejecutado en el servidor, generando HTML y enviándolo al cliente. El cliente recibirá los resultados de ejecutar el script, sin ninguna posibilidad de determinar qué código ha producido el resultado recibido. El servidor web puede ser incluso configurado para que procese todos los archivos HTML con PHP y entonces no hay manera que los usuarios puedan saber que tienes debajo de la manga.

Lo mejor de usar PHP es que es extremadamente simple para el principiante, pero a su vez, ofrece muchas características avanzadas para los programadores profesionales. Por ello se convierte en un lenguaje completo en el que resulta relativamente sencillo comenzar a implementar scripts simples e ir completándolos y mejorándolos según se adquieren conocimientos.

### 3.4 Web Analytics

El *análisis web* es un conjunto de técnicas relacionadas con el análisis de datos relativos al tráfico en un sitio web con el objetivo de entender su tráfico como punto de partida para optimizar diversos aspectos del mismo (24).

Existen dos categorías de análisis web: *off-site* y *on-site*. Los análisis web *off-site* hacen referencia a mediciones web y análisis independientes de si se es el dueño del sitio web o se le da mantenimiento. Esto incluye la medición de la audiencia potencial (oportunidad), participación de voz (visibilidad) y zumbidos (comentarios) acerca de lo que está pasando en Internet como un todo.

Los análisis web *on-site*, que será el que nosotros utilicemos, miden la ruta de un visitante una vez que se ingresa a un sitio de internet de propiedad de la persona que realiza el análisis. Esto incluye conversiones; por ejemplo qué páginas de llegada alienan a las personas a hacer una compra. Las mediciones de análisis web *on-site* comparan indicadores de desempeño clave y lo usan para mejorar un sitio web o la respuesta de la audiencia frente a una campaña de mercado.

Históricamente, el análisis web ha hecho referencia a la medición de visitantes *on-site*. Sin embargo, en años recientes se ha producido la convergencia de ambos aspectos, principalmente debido a que los vendedores están produciendo herramientas que abarquen ambas categorías.

Las herramientas de análisis web son las encargadas de capturar y procesar la información del sitio web, para proveer información sobre el comportamiento de los usuarios en el sitio: el sitio del que proceden, qué hacen en el sitio, por qué páginas navegan, durante cuánto tiempo, cuántas veces revisitan el sitio, de qué país son, qué tipo de conexión de internet tienen, en qué punto abandonan el sitio, en qué paso de un proceso de alta desisten, etc.

Las herramientas de Web Analytics pueden basarse en diferentes plataformas tecnológicas:

- ***Analizadores de ficheros de logs.*** Son programas que analizan los logs de los servidores proporcionando información sobre quién, cuándo y cómo los visita.

- ***Etiquetado de páginas.*** Este método, más reciente que el anterior, está basado en la incorporación de un script a cada una de las páginas de un sitio. Cada vez que una página es visitada, este script se comunica con una base de datos a la que comunica la impresión de la página junto con, potencialmente, datos adicionales procedentes de las cookies.
- ***Sistemas híbridos.*** Algunas empresas han desarrollado soluciones que combinan ambas soluciones agregándose las ventajas individuales de las mismas.

También existe un importante interés en cuanto a los factores económicos sobre todo desde el punto de vista del marketing, debido a que estas herramientas sirven para cuantificar los usuarios que visitan una web y por tanto pueden ver un anuncio hecho en dicha página, así como, determinan cuántos de estos usuarios acceden al enlace publicitario desde ahí y proporcionan datos de ellos como su ubicación, idioma, etc..

### **3.5 Otras herramientas**

#### **3.5.1 Editor de código fuente**

Para poder implementar la programación tanto de la interfaz como de la gestión de datos se necesita utilizar un editor de texto para código fuente, por lo que serviría cualquier programa que permita dicha edición, como cualquier editor que está instalado por defecto en los diferentes SO.

Un editor de código fuente es un editor de texto diseñado específicamente para editar el código fuente de programas informáticos. Puede ser una aplicación individual o estar incluido en un entorno de desarrollo integrado.

Los editores de código fuente tienen características diseñadas exclusivamente para simplificar y acelerar la escritura de código fuente, como resaltado de sintaxis, autocompletar pareo de llaves. Estos editores también proveen un modo conveniente de ejecutar un compilador, un intérprete, un depurador, o cualquier otro programa que sea relevante en el proceso de desarrollo de software. Por lo que, si bien muchos editores de texto pueden ser usados para editar código fuente sin problemas, si no mejoran, automatizan y facilitan la edición del código, no ameritan ser llamados "editores de código fuente", y son únicamente editores de texto que pueden ser usados para editar código fuente.

Algunos editores de código fuente verifican la sintaxis a medida que el programador escribe, alertando inmediatamente sobre los problemas de sintaxis que puedan surgir. Otros editores de código fuente comprimen el código, convirtiendo las palabras clave en tokens de un solo byte, eliminando espacios en blanco innecesarios y convirtiendo los números a una forma binaria. Estos editores tokenizadores descomprimen el código fuente al momento de visualizarlo, imprimiéndolo con los espacios y mayúsculas adecuadas. Existen editores que realizan ambas tareas.

### **3.5.2 Navegador web**

Es necesario estar en constante interacción con un navegador web, pues es en él donde podremos ir depurando la página e implementándola de cara al usuario final.

Un navegador o navegador web, es una aplicación que opera a través de Internet, interpretando la información de archivos y sitios web para que estos puedan ser leídos, ya se encuentre ésta alojada en un servidor dentro de la W3C o en un servidor local (25).

El navegador interpreta el código, HTML generalmente, en el que está escrita la página web y lo presenta en pantalla permitiendo al usuario interactuar con su contenido y navegar hacia otros lugares de la red mediante enlaces o hipervínculos.



La funcionalidad básica de un navegador web es permitir la visualización de documentos de texto, posiblemente con recursos multimedia incrustados. Los documentos pueden estar ubicados en la computadora en donde está el usuario, pero también pueden estar en cualquier otro dispositivo que esté conectado en la computadora del usuario o a través de Internet, y que tenga los recursos necesarios para la transmisión de los documentos.

Tales documentos, comúnmente denominados páginas web, poseen hipervínculos que enlazan una porción de texto o una imagen a otro documento, normalmente relacionado con el texto o la imagen.

El seguimiento de enlaces de una página a otra, ubicados en cualquier ordenador conectado a Internet, se llama navegación, de donde se origina el nombre navegador.

### **3.5.3 Programa de interpretación de datos.**

A la hora de poder hacer una interpretación de manera sistemática con todos los datos introducidos por los usuarios, necesitamos utilizar una herramienta de ámbito matemático/estadístico, donde obtener resultados que nos ayuden tanto cuantitativa como gráficamente a determinar la mejor predicción posible para nuestro estudio.

## **3.6 Selección y justificación para el desarrollo del PFC**

### **Hosting**

Los requisitos que a priori tuvimos en cuenta para el Hosting de nuestra página web fueron los siguientes:

- 100 MB de almacenamiento web
- Tráfico web que soporte al menos 100 conexiones simultáneas
- Un dominio web
- 2 ó 3 cuentas de correo para la gestión de la web
- Que soporte programación PHP
- 1 BBDD MySQL de 100 MB de capacidad
- Fiabilidad ante las posibles pérdidas de información

El Hosting elegido para poder alojar nuestra página web, fue el paquete LINUX Dual Básico de la empresa 1&1 ya que siendo el paquete más económico cubría con creces todas nuestras necesidades ofreciéndonos las siguientes características:

- 1 GB de almacenamiento web.
- Trafico web ilimitado.
- 1 Dominio incluido.
- 10 cuentas de correo con nuestro dominio.
- Soporta programación en PHP.
- 1 BBDD MySQL de 1 GB.
- Georredundancia de almacenamiento en dos centros de datos independientes. Sin caídas ni pérdidas.
- Editor web para 1 proyecto web.

A la vista de las necesidades finales de la web, llegamos a la conclusión de que la elección fue correcta debido a que las características en las cuales teníamos limitación por número o capacidad, cubrieron a la perfección dichas necesidades, siendo las siguientes:

- 5 MB de almacenamiento web.
- 1 dominio web.
- 1 cuenta de correo.
- 1 Base de datos de 500 KB.

El dominio elegido para implementar la web en internet será: ***petaso1x2.com***.

### **Entorno de desarrollo web**

Para poder obtener un entorno de desarrollo web completo necesitamos aunar tres requisitos fundamentales:

- Intérprete de PHP.
- Servidor HTTP Apache.
- Sistema de gestión de BBDD MySQL.

WampServer es un proyecto de código abierto y de uso gratuito (licencia GPL) desarrollado para Windows que constituye un entorno completo de desarrollo web de fácil uso con el que poder implementar de forma local una página web de manera prácticamente idéntica a como se va a implementar vía web. Este servicio ofrece lo siguiente:

- Apache 2.2.22.
- MySQL 5.5.24.
- PHP 5.3.13.
- PhpMyadmin 3.4.10.1.
- XDebug 2.1.2 .
- XDC 1,5.
- SQLBuddy 1.3.3.
- webGrind 1,0.

Por todas estas características, unido a su sencillo uso, decidimos utilizarlo en detrimento de otros entornos similares como el XAMPP.

## **Web Analytics**

Para poder tener un buen análisis sobre las estadísticas que conciernen a la página web con respecto a visitas, procedencia, duración, etc. necesitamos de un programa que nos haga un completo informe sobre muchos de estos aspectos para poder comprender cómo reaccionan los usuarios a diferentes maniobras de marketing. Por ello necesitamos a priori las siguientes características:

- Informes tanto diarios como por semanas/meses.
- Número de visitas globales y a páginas en concreto.
- Fuentes de tráfico.
- Datos demográficos y de idiomas.

La herramienta de *Google*, *Google Analytics* es un servicio gratuito de estadísticas de sitios web. Ofrece información agrupada según los intereses de tres tipos distintos de personas involucradas en el funcionamiento de una página: ejecutivos, técnicos de marketing y webmasters.

Se pueden obtener informes como el seguimiento de usuarios exclusivos, el rendimiento del segmento de usuarios, los resultados de la campaña de marketing, el marketing de motores de búsqueda, las pruebas de versión de anuncios, el rendimiento del contenido, el análisis de navegación, los objetivos y proceso de redireccionamiento o los parámetros de diseño web. Esta herramienta se desarrolló en base a la compra de Urchin (hasta entonces la mayor compañía de análisis estadístico de páginas web) por parte de *Google*.

Su funcionamiento es muy sencillo, se comienza añadiendo un código JavaScript a cada una de las páginas que se desea analizar, al que se denomina GATC (*Google Analytics Tracking Code*). Este carga algunos archivos desde los servidores *Google* y los monitoriza para luego enviar toda esta información al servidor *Google* y almacenarla en la cuenta de cada usuario.

*Google Analytics* ofrece muchos recursos formativos, tanto si se es experto como principiante, tiene numerosos vídeos y artículos de formación que ayudan a obtener el mayor rendimiento de esta herramienta. Tiene una interfaz muy completa de informes con gráficos desarrollados en Adobe Flash por lo que facilita aún más la comprensión de resultados.

Las principales funciones que nos ofrece son:

- Informes diarios y completamente configurables (se puede seleccionar cualquier rango de fechas).
- Informes en tiempo real (se pueden obtener datos de los visitantes conectados en tiempo real, aunque se encuentra en fase de pruebas).
- Datos demográficos (país, región y ciudad de los visitantes, así como el idioma de su PC).
- Datos de comportamiento (visitantes nuevos y recurrentes, así como el tiempo que perduran en la web).
- Tecnologías (navegador y SO que utilizan, así como la red de telefonía a la que pertenecen).
- Móvil (usuarios que utilizan móvil para conectarse, así como los dispositivos de los que se trata).
- Fuentes de tráfico (desde dónde accedieron a la web).
- Estadísticas por páginas de la web.
- Flujo de visitantes (esquema con cualquiera de las anteriores estadísticas en el que nos indica como navegan por las diferentes páginas de la web y cuando abandonan esta).
- Posibilidad de exportar estadísticas a correo electrónico o a varios tipos de archivos (CSV, PDF, etc.).
- Decenas de maneras de personalización.

Por todas estas características creímos idóneo el uso de *Google Analytics* para obtener las estadísticas de uso de la web.

### **Editor de código fuente**

Para poder implementar la programación tanto de la interfaz como de la gestión de datos se podría utilizar cualquier editor que está instalado por defecto en cualquier SO, pero sin embargo estos tienen grandes carencias que otros editores cubren sobradamente.

*NotePad++* es un editor de programación de código libre para diversos tipos de lenguaje (HTML, PHP, CSS, etc.), que se utiliza como editor de texto. Este editor permite tener varios documentos abiertos por la utilización de pestañas (multi-documento), visualización en ventana doble (multi-vista), coloración sintáctica para cada lenguaje, coloración a elección del usuario, utilización de marcadores (puntos de marca), soporte para arrastrar/colocar, soporte buscar/reemplazar, autocompletado, herramienta de zoom, etc. lo que facilita enormemente la tarea del programador, pues es mucho más sencillo observar y subsanar los errores, tanto de manera inmediata como mediante sus herramientas de búsqueda.

Por todo ello, y destacando su facilidad de uso, de comprensión y su detección de errores automática que facilita enormemente la labor del desarrollo decidimos utilizarlo para implementar todo el código del proyecto.

### **Navegador web**

A la hora de poder desarrollar de una manera aceptable para todo tipo de usuarios, hay que tener en cuenta la variedad de navegadores que se usan, incluidas las plataformas móviles (como podremos comprobar en el apartado 3 del Anexo B, estos dispositivos representan un significativo 14'7% de las conexiones web que recibimos), ya que todos interpretan de una manera parecida el código HTML y CSS pero no de forma idéntica si no se tienen determinadas precauciones, es decir, lo que en un navegador se ve bien, en otro se puede ver de manera incorrecta. Por ello es necesario ir desarrollando la página web e ir comprobando en paralelo las visualizaciones de éstos y subsanando los errores que van apareciendo.

Otra razón muy a tener en cuenta es que muchos de estos navegadores tienen herramientas exclusivas de desarrollo de páginas web para los desarrolladores que facilitan de una manera muy sustancial la implementación, colocación y decoración de cada uno de los elementos insertados en la web, así como por supuesto también a la hora de subsanar errores.

Por todo ello para el desarrollo de nuestra página web hemos utilizado los siguientes navegadores por volumen de usuarios según datos obtenidos con *Google Analytics* incluidos en el apartado 2 del Anexo B:

- *Google Chrome.*
- *Internet Explorer.*
- *Safari.*
- *Mozilla Firefox.*
- *Android Browser.*

De manera adicional y como se ha comentado antes, se han utilizado herramientas para desarrolladores de *Google Chrome* y *Mozilla Firefox*.

### **Interpretación de datos**

*MATLAB* (abreviatura de *MATrix LABoratory*, "laboratorio de matrices") es un software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio, lenguaje M. Está disponible para las plataformas Unix, Windows y Mac OS X.

Entre sus prestaciones básicas se hallan: la manipulación de matrices, la representación de datos y funciones, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware.

Adicionalmente nos parece muy interesante la posibilidad que aporta esta herramienta, pues se pueden establecer fácilmente comparativas gráficas entre los diferentes resultados y así obtener una rápida idea visual de la magnitud de los resultados.

Es un software muy usado en universidades y centros de investigación y desarrollo, como por supuesto en la *UC3M* dónde es el programa más utilizado por los estudiantes de Telecomunicaciones para la mayoría de sus asignaturas.

Por estos motivos y debido a que tenemos experiencia con dicha herramienta, será la que utilicemos para poder tratar los datos obtenidos a lo largo de las jornadas con el fin de obtener de manea cuantitativa y gráfica resultados que podamos interpretar para ir mejorando el algoritmo y así poder ir rectificando los errores con el fin de hallar el resultado más idóneo posible.

### **3.7 Uso de las herramientas de desarrollo web**

Los entornos de desarrollo web, tanto local como final, nos supusieron una pérdida de tiempo importante debido a nuestra inexperiencia. Por este motivo creemos conveniente explicar brevemente el funcionamiento de dichas herramientas.

#### **3.7.1 WAMP**

El uso de WAMP es relativamente sencillo siempre que se tengan unos conocimientos básicos sobre él, ya que si no puede plantear grandes problemas de comprensión y de uso.

El primer paso para utilizar el WAMP es su instalación. Se trata de un ejecutable de sencillos pasos en los que sólo cabe destacar la elección del ejecutable del navegador por defecto que queramos utilizar cuando nos dé la opción, ya que si no por defecto será el Internet Explorer y la configuración del usuario que se ha de hacer. Esta configuración consiste en la definición del servidor, usuario y contraseña. Estos datos son los que posteriormente se usarán en los scripts de interacción con las BBDD de MySQL por lo que es importante recordarlos. Debido a que usaremos una configura-



ción local y no se necesitan medidas de seguridad, los datos a introducir que recomendamos son los siguientes:

- **Servidor:** “localhost”.
- **Username:** “webmaster@localhost”.
- **Password:** (vacío).

Una vez configurado se creará un directorio con la carpeta WAMP y dentro de esta la carpeta *WWW*. Dicha carpeta adquiere suma importancia, pues será la carpeta raíz de nuestro entorno de desarrollo donde debemos introducir todos y cada uno de nuestros ficheros con los scripts pertinentes y desde donde se ejecutarán.

Para que nuestro ordenador funcione como un servidor local, hemos de arrancar el programa. Tras ello nos aparecerá el icono de la herramienta en la barra de notificaciones. Se puede configurar el idioma con el botón derecho del ratón, en este caso lo ponemos en español, y tras esto desplegamos el menú con el botón izquierdo como aparece en la *Figura 3*:



*Figura 3: Menú WAMP.*

Como se puede observar el símbolo de WAMP de la barra de notificaciones es de color verde, esto significa que todos los servicios funcionan correctamente. Si estuviera en rojo significa que el servicio está detenido, por lo que haría falta iniciarlo con la opción pertinente del menú. Si estuviera en naranja significa que el servicio está iniciado, pero hay algún problema con alguno de los componentes, por lo que habría que reiniciar el servicio, y si el sistema persiste, actualizar las versiones del componente que da error o volver a reinstalar el sistema.

En dicho menú se podrán ver las diferentes versiones de los elementos que componen la herramienta, como son Apache, PHP y MySQL. Así mismo aparecen los tres elementos que necesitamos:

- **Directorio WWW.** Es la carpeta almacenadora del directorio raíz, con clicar en esta opción se nos abrirá la ruta de dicha carpeta, la cual debe de contener todos los scripts de la página.
- **Localhost.** Es el directorio raíz, si pinchamos sobre esta opción, se ejecutará el navegador predeterminado que tengamos configurado con la ruta: localhost/index.php, donde en principio habrá una página de muestra perteneciente a la herramienta. Nuestra página de inicio ha de sustituir a esta para que se ejecute al iniciar la ruta. Por tanto o bien creamos nuestra página de inicio directamente con el nombre index.php, o creamos un script con dicho nombre en el que nos redirija a la página de inicio deseada.
- **PhpMyAdmin.** Es la aplicación web de interacción con las BBDD. Se ejecuta el navegador con la ruta: localhost/ phpMyAdmin, donde se encuentra dicha interfaz. El primer paso a realizar es la creación de una nueva BDD cuyo nombre utilizaremos en los scripts de nuestra web para acceder a ella. Es sumamente importante dar los privilegios correspondientes a esta base de datos, en este caso todos, por lo que una vez seleccionada nuestra BDD en la pestaña privilegios seleccionamos todo y aplicamos la configuración, ya que sin ello no podremos acceder a dicha BDD desde nuestra web. A partir de aquí simplemente se trata de crear y modificar tablas de datos con nuestro

criterio, teniendo en cuenta los parámetros que posteriormente vamos a necesitar. Dicha interfaz tiene muchas posibilidades, entre ellas cabe destacar la importación/exportación de tablas y BBDD con los que podremos fácilmente trasladarlas a otras aplicaciones de gestión de BBDD.

Con estas sencillas instrucciones no debería suponer mayor problema el manejo de la herramienta, aunque siempre van surgiendo dudas, que se subsanan con la experiencia de su uso.

### 3.7.2 Panel de control 1&1

Al igual que el entorno de WAMP, el uso de este panel de control es sencillo, de hecho en mayoría es bastante parecido al anterior.

Una vez se ha contratado el hosting pertinente, se puede acceder a dicho panel de control en el que se tiene asignada una BDD, un dominio y un subdominio de pruebas.

A continuación exponemos las principales funcionalidades de dicha herramienta:

- ***Dominios.*** En dicho panel, encontramos un enlace a “dominios” donde podemos administrar a qué ruta de enlace vinculamos cada uno de los dominios que tengamos. Por defecto tendremos nuestro propio dominio, y un subdominio de pruebas que nos facilita el proveedor para poder lanzar nuestra web a internet sin que nadie pueda acceder a ella sin la ruta de enlace. Simplemente se trata de elegir la ruta a la que vinculamos esos dominios para “abrir” nuestra web al público o no.
- ***Gestión de MySQL.*** Se trata simplemente de una interfaz de características similares a phpMyAdmin explicado anteriormente, donde se gestiona la BDD seleccionada.

- ***1&1 Gestión de Archivos.*** Es básicamente la carpeta virtual almacenadora de todos los archivos y scripts de nuestra web, a donde vincularemos los enlaces de nuestros dominios. Es equivalente al directorio WWW de WAMP.
  
- ***Cuentas de e-mail y 1&1 WebMail.*** Son la herramientas de creación/gestión de cuentas de correo electrónico vinculadas al dominio contratado y su administración.

Adicionalmente existen decenas de otras herramientas de gestión de usuario, gestión de facturación, comunicación, etc. las cuales son más o menos útiles dependiendo del uso que vayamos a dar a la web.

# CAPÍTULO 4

## DISEÑO Y MEDIDAS DE CALIDAD DEL SISTEMA

### 4.1 Tipos de “sites” de quinielas: información y procesado

La mayoría de webs de peñas quinielísticas o simplemente dedicadas a la predicción y apuestas de La Quiniela, basan su criterio a la hora de hacer sus pronósticos en datos históricos, en los datos ya introducidos por otros usuarios y en otros datos como por ejemplo las cuotas que dan las principales casas de apuestas sobre el posible resultado de los partidos.

De esta forma, cuando un usuario se dispone a hacer su predicción, se ve influenciado por una gran cantidad de datos que recibe de una manera directa, ya que en la mayoría de estas webs la información suele ir en el mismo boleto donde se introduce la predicción o en su defecto en tablas contiguas. Dicha información no es sólo objetiva, lo que podría dar simplemente información adicional al usuario que puede que desconociera, si no que hay (y parece ser a la que dan más importancia pues es la que ocupa un lugar prioritario de la web) una gran cantidad de información subjetiva como es el porcentaje de predicciones realizadas a cada una de las posibles opciones por los usuarios que ya han participado.

Cabe destacar que en todas estas webs, los usuarios participan en la quiniela con su propio dinero, es decir, han de abrir una cuenta, o introducir dinero en una “bolsa” con la que van pagando las diferentes apuestas. A parte del coste que conlleva realizar las apuestas pertinentes, dichas webs suelen cobrar una comisión por cada apuesta realizada que es la base de sus ingresos, pues normalmente suelen cobrar el 20% de la apuesta realizada a razón de gastos de gestión.

La predicción se realiza a través del método tradicional de apuesta en La Quiniela, es decir, intentar adivinar qué equipo será el ganador del partido o si acabarán en tablas, con la manera tradicional del 1X2. El 1 representa la victoria del equipo local, la X el empate y el 2 la victoria del equipo visitante.

Algunos links de estos tipos de webs son:

- [Quinielista.com](http://Quinielista.com).
- [Elnogmo1x2.com](http://Elnogmo1x2.com).
- [Quinielaonline.com](http://Quinielaonline.com).

## **4.2 Diseño propuesto: interfaz y procesado básico**

Como se ha explicado anteriormente, la mayoría de páginas web relacionadas con las apuestas de La Quiniela, ofrecen a sus usuarios cantidad de información tanto objetiva como subjetiva, que más que probablemente influenciará de manera notable en las decisiones que tomen a posteriori dichos usuarios. Basándonos en la filosofía que vamos a utilizar descrita en el capítulo 2, creemos que para poder obtener una predicción lo más idónea posible, hemos de conseguir que los usuarios tomen sus propias decisiones de manera independiente sin influencia alguna del grupo. No obstante sí que ponemos a disposición del usuario en el apartado “información útil”, las tablas clasificatorias de primera y segunda división de la LFP para que en el caso que lo deseen, dispongan de datos sobre qué resultados han ido obteniendo en la presente temporada. En ningún caso, la información es mostrada en la página principal o junto al boleto de predicciones, y mucho menos el resultado parcial obtenido con la participación de los demás usuarios, pues creemos que es indispensable la objetividad y la no influencia en las decisiones de cada individuo.

Para poder recoger la máxima cantidad de información posible de los usuarios, así como lograr una amplia participación, llegamos a la conclusión de que debíamos cambiar el modelo de recolección de datos empleado por la mayoría de las páginas web basadas en la predicción de quinielas.

Dicho modelo, como se ha explicado anteriormente consta de la simple predicción sobre qué equipo será el ganador del partido o bien si este acabará en tablas.

Para ampliar cuantitativamente esta recogida de información planteamos un sistema de predicción basado en diez posibilidades en lugar de tres. Una de las posibilidades es la de total incertidumbre, es decir, creer que cualquier resultado es posible con la misma probabilidad. Las otras nueve posibilidades se agrupan de tres en tres de tal forma que cada grupo representa uno de los tres posibles resultados: victoria local, empate y victoria visitante. Dentro de esos grupos, cada uno de los tres resultados posibles implica un grado diferente de confianza en cómo se va a cumplir dicho resultado, es decir, variará desde el pleno convencimiento de que se produzca, a la creencia de que se producirá pero por una diferencia muy ajustada en el partido.

Para poder recabar dicha información optamos por un modelo de introducción de datos similar al de los demás web site de quinielas, donde se presenta emulando el clásico boleto de apuestas, eso sí, adecuado a nuestro nuevo criterio de cuantización de los pronósticos. Como se puede apreciar en la *Figura 4* se trata de las quince filas con los diferentes partidos que componen la jornada de La Quiniela y sus diez opciones, donde el usuario podrá seleccionar una y sólo una para cada partido.

Envíanos tu predicción											
		0	1	1	1	x	X	x	2	2	2
1	Valladolid - Espanyol	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Atlético - Málaga	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3	Real Betis - R. Sociedad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4	Athletic - Osasuna	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5	Levante - Valencia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6	Zaragoza - Getafe	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7	Mallorca - Granada	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8	Rayo - Deportivo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9	Ponferradina - Sabadell	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10	Las Palmas - Almería	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11	Sporting - Villarreal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12	Huesca - Girona	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13	Recreativo - Alcorcón	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14	Murcia - Racing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15	Barcelona - R. Madrid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**ENVIAR**

Figura 4: Ejemplo de boleto de predicción.

Para que el usuario comprenda de una manera aproximada lo que pretendemos que se entienda por cada uno de los diez niveles de pronóstico, junto al boleto incluimos unas sencillas instrucciones de uso:

- [ 0 ] - No sabría decir ningún resultado.
- [ 1 ] - Ganará seguro el equipo local.
- [ 2 ] - Ganará probablemente el equipo local.
- [ 3 ] - Ganará el equipo local pero muy ajustado.
- [ 4 ] - Empate pero más opciones del equipo local.
- [ 5 ] - Empatarán, será un partido igualado.
- [ 6 ] - Empate pero más opciones del equipo visitante.
- [ 7 ] - Ganará el equipo visitante pero muy ajustado.
- [ 8 ] - Ganará probablemente el equipo visitante.
- [ 9 ] - Ganará seguro el equipo visitante.



Estos diez niveles los representamos del 0 al 9, para posteriormente poder tratar estos datos, pero en realidad, para hacerlo de una manera visualmente más sencilla para el usuario, cambiamos los números del 1 al 9 y los sustituimos por 1, X y 2, de mayor y menor tamaño representando la confianza que se tiene en dicho resultado como se habrá podido apreciar con anterioridad en la *Figura 4*.

Una vez obtenida la previsión sobre los diferentes partidos de los usuarios que han decidido participar en el juego, hemos de someter estos datos a un análisis para obtener un único resultado el cual será nuestra predicción y por la que apostaremos en el boleto de cada jornada de La Quiniela.

Como un primer análisis básico de los datos, decidimos que la forma en la que se decidirá el resultado final, sea la media ponderada de las estimaciones independientes e individuales de las predicciones de los usuarios basándonos en el anteriormente explicado método de inteligencia colectiva estadística. Por tanto, sin tener en cuenta las predicciones que dieron valor 0, ya que no son válidas en este sistema, se efectúa una media ponderada entre los resultados del 1 al 9 en cada partido teniendo en cuenta el número de usuarios que sí efectuó un voto válido, es decir, distinto de 0.

Una vez obtenida dicha media, y puesto que los posibles valores varían del 1 al 9, posicionamos dos fronteras de decisión tal que los tres espacios de decisión sean equivalentes, es decir, que tengan igual longitud. Dicha longitud vendrá dada por:  $(9-1)/3 = 2,66$ . Por tanto nuestras fronteras de decisión se situarán en:  $F_{1,X} = 1 + 2,66 = 3,66$  y  $F_{X,2} = 9 - 2,66 = 6,33$ .

Es decir, los espacios de decisión de las tres posibilidades quedarán de la siguiente manera:

- $E_1 = [1 \sim 3,66)$ .
- $E_X = [3,66 \sim 6,33]$ .
- $E_2 = (6,33 \sim 9]$ .

De manera adicional, hemos tenido en cuenta si la predicción la efectuaba un usuario logueado o un usuario anónimo, pues podría darse el caso de que tuviéramos una gran participación de usuarios anónimos los cuales podrían querer desestabilizar el resultado, por ello y para minimizar daños, contabilizamos sus predicciones con el 50% del peso de los demás usuarios y hacemos una comparativa entre la media ponderada de todos los usuarios y la media ponderada contabilizando a los usuarios anónimos al 50% como podemos observar en la siguiente tabla de ejemplo perteneciente a una jornada real.

Nº de participantes	Nº de anónimos	Nº con anónimos ponderado
21	1	20.5

Nº	Local	Visitante	Total Pétaso	Media Pétaso	Decision Pétaso	Total con Anónimos Ponerados	Media con Anónimos Ponerados	Decision Anónimos Ponerados
1	Atlético	Valladolid	31	1.48	1	30.5	1.49	1
2	Real Betis	Espanyol	79	3.76	X	78	3.8	X
3	Athletic	Málaga	106	5.3	X	103.5	5.31	X
4	Levante	R.Sociedad	95	4.52	X	93.5	4.56	X
5	Zaragoza	Osasuna	83	3.95	X	82.5	4.02	X
6	Celta	Getafe	100	4.76	X	99.5	4.85	X
7	Barcelona	Granada	27	1.29	1	26.5	1.29	1
8	Rayo	R.Madrid	163	7.76	2	158.5	7.73	2
9	Sabadell	Xerez	102	5.1	X	99.5	5.1	X
10	Las Palmas	Villarreal	145	7.25	2	141.5	7.26	2
11	Córdoba	Girona	74	3.7	X	73.5	3.77	X
12	Sporting	Alcorcón	72	3.6	1	71.5	3.67	X
13	Huesca	Elche	114	5.7	X	113.5	5.82	X
14	Numancia	Murcia	99	4.95	X	98.5	5.05	X
15	Mallorca	Valencia	120	6	X	117.5	6.03	X

*Figura 5: Ejemplo de resultados de una jornada.*

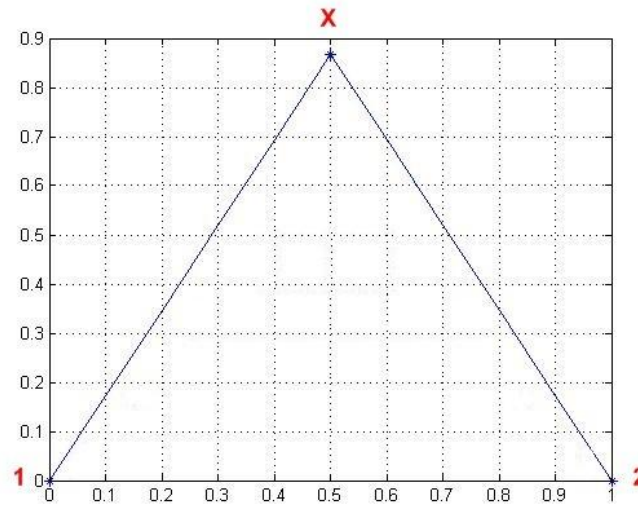
Como se puede observar en la *Figura 5*, en la jornada a la que pertenecen los datos participó un usuario anónimo, por lo que el resultado ponderando los anónimos varió con respecto al resultado de media ponderada simple. Cabe destacar que en dicha jornada, la influencia del mencionado usuario anónimo tuvo relevancia, pues en la casilla número 12 variaba el resultado del decisor dependiendo de si dicho usuario se contabilizaba o no como usuario anónimo siguiendo el criterio explicado anteriormente.

### 4.3 Métricas de calidad: compromiso entre usabilidad e información

Estamos convencidos de que podríamos mejorar nuestros resultados notablemente si pudiéramos aplicar una cuantización aun mayor en nuestra recogida de datos y que ésta fuera lo más parecida posible al pensamiento original del usuario. Esto sería factible dando plena libertad en el reparto de probabilidades dentro de los posibles resultados, pero a cambio crearíamos un rechazo aun mayor al usuario a la hora de participar ya que está demostrado que cuanto más simple es un juego, más participación genera, y por tanto, a mayor complejidad, más reticentes son los usuarios a colaborar.

Como hemos comentado, creemos que lo ideal para el desarrollo idóneo de la predicción sería poder ofrecer una constelación completa de posibilidades entre las tres posibles opciones. Éste quedaría visualmente representado por un triángulo en el que los vértices son las tres posibilidades (1, X y 2), y sus lados tienen una longitud de 1, refiriéndonos métricamente a términos de probabilidad de que ocurra un suceso determinado. El área de dicho triángulo cubre la totalidad de las probabilidades de decisión, siendo la distancia del punto elegido a cada uno de los vértices inversamente proporcional a la probabilidad de que ocurra cada uno de dichos sucesos, es decir, en el triángulo donde las probabilidades son  $[P_1, P_X, P_2]$ , un punto situado justo en el vértice de la X daría una probabilidad de  $[0, 1, 0]$  y un punto situado en el centro de gravedad del triángulo daría una probabilidad de  $[0.333, 0.333, 0.333]$ .

Para poder entender más fácilmente el mencionado triángulo de probabilidades vamos a representarlo visualmente en *la Figura 6*:



*Figura 6: Triángulo de probabilidades  $[P_1, P_X, P_2]$ .*

Si implementáramos una interfaz gráfica donde se pudiera seleccionar a través del triángulo las probabilidades de los tres eventos a lo largo de todo el área del triángulo, y por tanto cubriendo con una cuantización prácticamente infinita las probabilidades de los tres eventos, resultaría algo complejo de entender para la mayoría de los usuarios, por lo que creemos sería un obstáculo a la hora de conseguir la mayor participación posible, pues pensamos que cuanto más sencillo sea la introducción de pronósticos, mayor será el volumen de usuarios estables que participen en nuestra web.

# CAPÍTULO 5

## PROCESADO DE DATOS, EVALUACIÓN Y MODIFICACIÓN DEL SISTEMA.

### 5.1 Descripción de los datos recogidos

El problema de la recogida de datos que efectúa nuestra interfaz es que crea una base de conocimiento unidimensional, es decir, representamos la probabilidad de que pase uno de los tres sucesos desestimando los demás como un conjunto sobrante. Lo que hemos de intentar es tratar de hallar la manera de transformar nuestros datos unidimensionales en datos bidimensionales que cuantifiquen las probabilidades de los tres sucesos posibles.

Para poder lograr nuestro objetivo, llegamos a la conclusión de que el planteamiento idóneo sería apoyarnos en la teoría de un sistema de decisiones Bayesiano en el que, como explicamos en el capítulo 2, asignando valores numéricos a diferentes grados de creencia sobre un resultado concreto y haciendo que estos respondan a razonamientos del sentido común, podremos utilizarlos como reglas de probabilidad de suma y/o producto.

Por dicho motivo, decidimos asignar cada una de las posibles opciones a probabilidades repartidas entre los tres resultados globales 1, X y 2 dadas por los diferentes grados de confianza que se pretenden desprender de los criterios asignados anteriormente.

La siguiente *Figura 7* muestra las probabilidades, en porcentaje, de las diez opciones de las que disponen los usuarios para poder expresar su pronóstico:

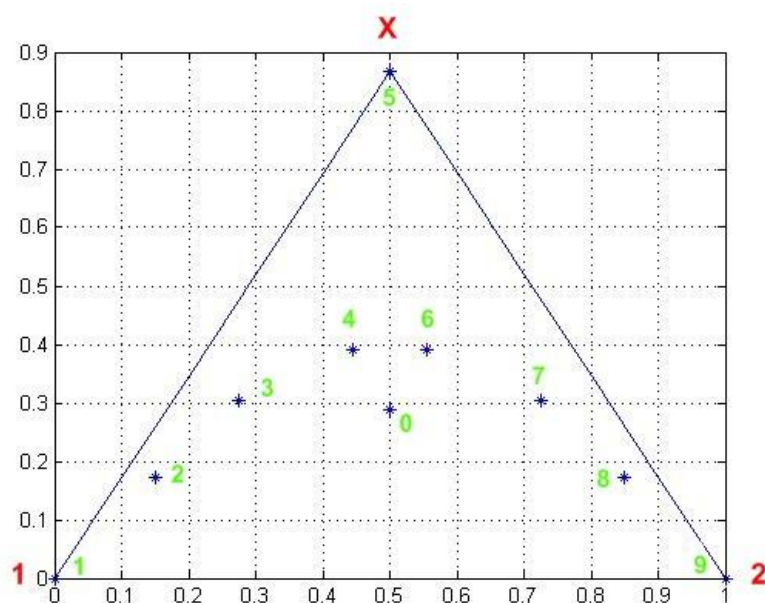
	1	X	2
0	33,3	33,3	33,3
1	100	0	0
2	75	20	5
3	55	35	10
4	33	45	22
5	0	100	0
6	22	45	33
7	10	35	55
8	5	20	75
9	0	0	100

*Figura 7: Probabilidades asignadas a los tres resultados según la predicción.*

Tras deliberar sobre la opción de mostrar a los usuarios alguna interfaz gráfica con los datos de la tabla anterior, llegamos a la conclusión de que no sería conveniente explicarles lo que pretendíamos a través de números y porcentajes, pues supone un aumento de la complejidad en la toma de decisiones y por tanto corríamos el riesgo de que la participación bajara drásticamente. Por otro lado entendemos que a través de las sencillas explicaciones que se muestran como norma de predicción, estamos expuestos a una interpretación diferente por parte de los usuarios que la que inicialmente pretendemos mostrar, pero al menos en una fase inicial creemos más conveniente tener una población de usuarios lo más amplia posible.

Como podemos observar en la tabla anterior, para cada grado de confianza en un resultado, aplicamos probabilidades a cada uno de los tres resultados. Poniendo un ejemplo, si elegimos la opción 7: [10, 35, 55], estamos expresando nuestra convicción de que el equipo visitante tiene algo más de la mitad de las posibilidades de ganar, el empate es bastante posible, y la victoria local es un resultado poco probable.

Con esto pretendemos convertir la inicial recogida de datos unidimensional, en una bidimensional, que perfectamente podía estar representada en el triángulo de probabilidades anteriormente expuesto. A continuación mostramos la constelación formada por los diez puntos elegidos representados en dicho triángulo en la *Figura 8*.



*Figura 8: Constelación de puntos de decisión.*

Como se puede observar, hay un cierto vacío en la parte inferior del triángulo que representaría la probabilidad de que un partido acabara en empate fuera muy baja pero sin embargo la probabilidad de que ganara el equipo local o el visitante fuera muy alta, en torno a los rangos  $[0.3 \sim 0.7, 0 \sim 0.2, 0.3 \sim 0.7]$ . Desgraciadamente es una probabilidad que con nuestro criterio de una sola dimensión no podemos aplicar, pero sería un paso adelante para la mejora del algoritmo, aunque entendemos que hemos de sacrificar cierta recogida de información con tal de atraer al mayor número de participantes posible.

En la *Figura 9* que mostramos a continuación, podemos ver la tabla resumen de los resultados obtenidos en una jornada, donde se pueden observar los porcentajes ponderados de cada partido para las tres posibles opciones y la elección se basa en el de mayor probabilidad.

Nº de participantes						
21						
Nº	Local	Visitante	% 1	% X	% 2	Decisión
1	Atlético	Valladolid	88.3	9.3	2.4	1
2	Real Betis	Espanyol	45.1	36.9	18	1
3	Athletic	Málaga	28.1	38.3	33.6	X
4	Levante	R.Sociedad	38.7	34	27.3	1
5	Zaragoza	Osasuna	39	46.7	14.3	X
6	Celta	Getafe	28.7	48.1	23.2	X
7	Barcelona	Granada	93.3	5.2	1.4	1
8	Rayo	R.Madrid	10.6	11.9	77.5	2
9	Sabadell	Xerez	19.4	60.2	20.4	X
10	Las Palmas	Villarreal	10.8	29	60.2	2
11	Córdoba	Girona	44.7	37.3	18	1
12	Sporting	Alcorcón	47	37.3	15.7	1
13	Huesca	Elche	24.1	36.1	39.8	2
14	Numancia	Murcia	26.3	50.4	23.3	X
15	Mallorca	Valencia	15	46.8	38.2	X

*Figura 9: Ejemplo de probabilidades resultantes en una jornada.*



## 5.2 Evaluación de los datos: el papel de la interfaz

Creemos que a la hora de poder recabar información sobre los pronósticos predichos por los usuarios en un proyecto como el nuestro, la interfaz juega un papel importantísimo en la obtención de datos o en nuestro caso predicciones. Es importante que dicha interfaz represente y pueda cumplir lo más fielmente posible los objetivos para los que fue diseñado, es decir, básicamente lograr una participación lo más amplia posible y que los pronósticos obtenidos por dicha participación reflejen exactamente el pensamiento de cada uno de los usuarios sobre los diferentes encuentros.

Para poder cumplir las anteriores condiciones creemos indispensables los siguientes requisitos:

- ***Buena estética:*** es imprescindible crear una interfaz llamativa que invite a la participación, aunque todo en su justa medida, es preferible algo sencillo y bien estructurado que algo demasiado recargado y llamativo.
- ***Sencillez:*** ante todo, es fundamental que la interfaz sea lo más sencilla posible a la hora de recoger la información del usuario, pues mientras más dudas tenga éste al cumplimentarlo más rechazo creará.
- ***Distribución de la web:*** un aspecto muy importante es la facilidad de visualización de todos y cada uno de los contenidos de la web, es decir, que el usuario pueda navegar sencillamente, ya que mientras menos dificultades genere la navegación, menos rechazo creará a abandonar la web.
- ***Pronósticos entendibles:*** a la hora de recabar información sobre los pronósticos que cada usuario prevé, es fundamental que estos queden perfectamente definidos y que así dicho usuario pueda identificarlo con alguno de los propuestos.

- ***Pronósticos que cubran todo el rango de probabilidades:*** una de las principales necesidades de la interfaz es crear una constelación de posibilidades de pronóstico que cubra en la mayor medida posible todo el área del triángulo de probabilidades descrito anteriormente.

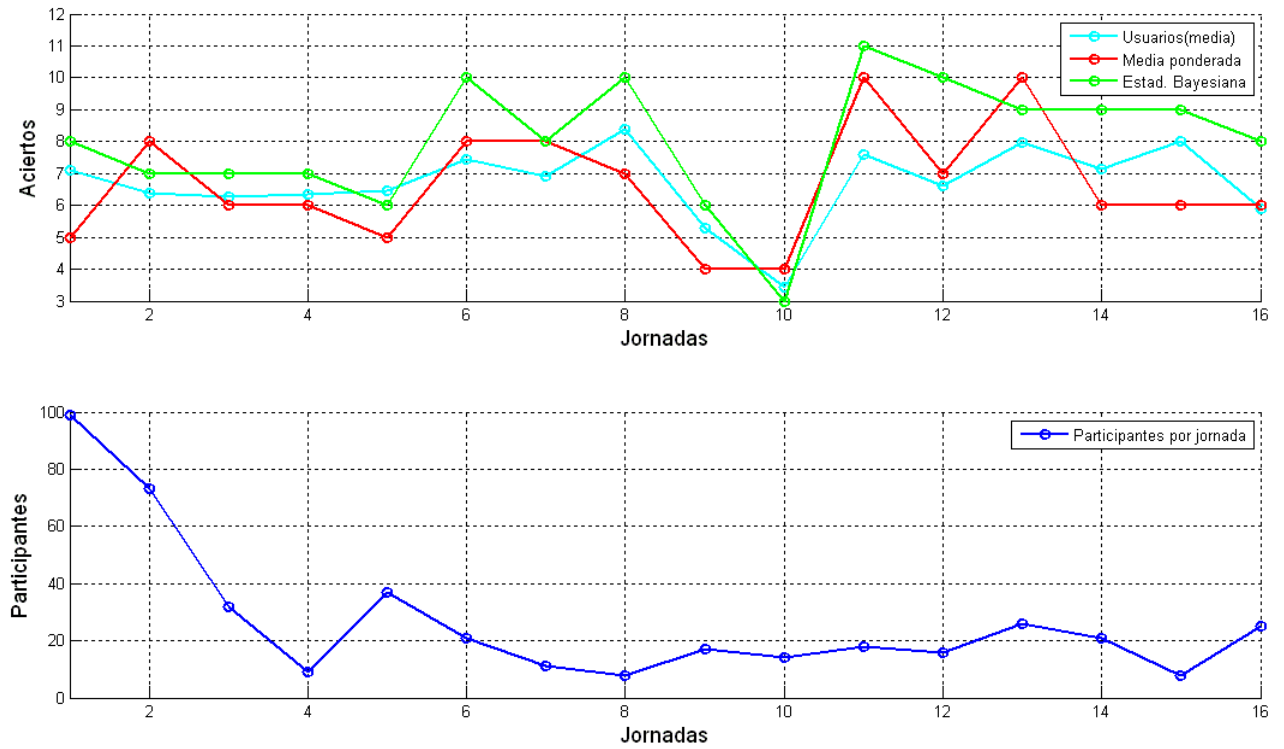
Estos criterios desgraciadamente no se pueden cumplir conjuntamente en su máxima expresión, pues una interfaz que recoja de la manera más óptima y amplia posible los pronósticos de los usuarios, será compleja y por lo tanto, creará rechazo.

Por ello, creemos que en el caso de ser posible, lo ideal sería poder implementar una que aunara la gran capacidad de recolección de información del triángulo de probabilidades  $P_T = [P_1, P_X, P_2]$  descrito en el capítulo 4, con la sencillez del formato de la constelación de 10 puntos que estamos utilizando.

## **5.3 Análisis de los datos obtenidos**

### **5.3.1 Resultados de aciertos y participación**

A lo largo de todas las jornadas que hemos podido obtener pronósticos de los usuarios, pertenecientes a finales de la temporada 2011/2012 y comienzos de la 2012/2013. Estos datos fueron evaluados y obtuvimos pronósticos en base a la evolución de mejora de nuestro algoritmo de decisión colectiva, no obstante, para poder realizar una comparativa entre los dos métodos utilizados, empleamos dichos predictores en las diferentes jornadas y los comparamos con el acierto medio de los usuarios, resultando la gráfica de la *Figura 10*:



*Figura 10: Aciertos obtenidos y participación de usuarios en cada jornada.*

Queremos recalcar que los resultados mostrados por la gráfica, pertenecen al total de aciertos sobre las 15 casillas, sin diferenciar el pleno al quince del resto como sí se hace a la hora del reparto de premios.

Como podemos observar, por norma general el algoritmo de la matriz de probabilidades basado en la decisión bayesiana, tiene un mejor resultado que el de la media ponderada de los pronósticos, obteniendo de media 1'375 aciertos más por jornada.

Queda demostrado que se consigue mejorar el sesgo que nos influía como se explicó en el capítulo anterior al efectuar la simple media ponderada de los pronósticos de los usuarios en nuestra interfaz.

Una propiedad relacionada con el sesgo es la de la consistencia: un estimador puede tener un sesgo pero el tamaño de éste converge a cero conforme crece el tamaño muestral. En principio, y a la vista de los resultados se podría afirmar que nuestro

estimador es inconsistente debido a que no obtuvimos mejores resultados cuando contamos con una mayor población de usuarios en dicho estimador, aunque estos datos no son del todo fiables ya que la continua mejora de la interfaz pudo influir en la mejor obtención de pronósticos.

Un dato significativo que se refleja claramente en la gráfica, es la mejora de los resultados una vez implementamos la opción de pronóstico 0 de máxima incertidumbre (misma probabilidad para los tres eventos posibles: 1, X y 2), que se implantó de la sexta jornada en adelante. Como se aprecia, los resultados mejoraron significativamente, sobre todo con el sistema de la matriz de probabilidades, donde se obtuvo de media 8'45 aciertos por jornada, y si no tenemos en cuenta las jornadas 9 y 10, se obtuvo 9'33 aciertos por jornada en media. Estas dos jornadas que apartamos del segundo cálculo se pueden considerar prácticamente *outsiders*, ya que como se deduce del escrutinio de ambas jornadas por parte de LAE, hubo varios resultados que se produjeron contra todo pronóstico, por lo que incluso varios premios de La Quiniela quedaron desiertos.

Un dato plenamente objetivo, es el vertido por los resultados del escrutinio de algunas de las jornadas. Desgraciadamente no es dato que podamos aplicar a todas las jornadas ya que dicho escrutinio sólo publica los acertantes de 10 combinaciones en adelante, por lo que podremos evaluar únicamente las jornadas en las que tuvimos 9 y 10 aciertos sin contar con el pleno al quince. Desde que optimizamos la interfaz y aplicando nuestro decisor más elaborado, han transcurrido once jornadas, de las cuales seis cumplen este requisito, otras tres están justo un peldaño por debajo, es decir en los ocho aciertos, y las dos restantes son los *outsiders* comentados anteriormente condicionados por la dificultad de algunos pronósticos.

A continuación mostramos la tabla de la *Figura 11* que representa el porcentaje de apuestas que se realizaron y consiguieron obtener más aciertos que la nuestra.

Nº	Jornada LAE	Aciertos	Apuestas	Apuestas mejores	Porcentaje mejores
6	53 (11/12)	9 + pleno	18.846.334	559.942	2,97%
8	56 (11/12)	9 + pleno	7.937.950	132.317	1,67%
11	3 (12/13)	10 + pleno	13.411.303	284.679	2,12%
12	4 (12/13)	10	6.690.862	21.816	0,33%
13	5 (12/13)	9	17.460.671	3.406.892	19,51%
14	7 (12/13)	9	15.731.333	2.627.139	16,70%

*Figura 11: Resumen acertantes con 10 o más aciertos.*

Como se muestra en la tabla los resultados obtenidos son realmente positivos, ya que en cuatro de las ocasiones no nos llegó a superar el 3% de las apuestas realizadas y en las otras dos no llegó al 20%.

En el caso de las jornadas número 3 y 4 de la presente temporada, podemos obtener aún más datos, pues obtuvimos 10 aciertos sin contar el pleno al quince, por lo que sabemos exactamente el porcentaje de apuestas que fueron inferior al nuestro al poder contabilizar las apuestas que obtuvieron los mismos aciertos que nosotros.

Como se puede comprobar en la tabla de la *Figura 12*, en la jornada 3 y 4 de la presente campaña obtuvimos mejores resultados que aproximadamente el 92% y el 98% de las apuestas totales respectivamente. Lo que demuestra un alto éxito en esas dos jornadas.

Nº	Jornada LAE	Aciertos	Apuestas	Apuestas peores	Porcentaje peores
11	3 (12/13)	10 + pleno	13.411.303	12.336.593	91,99%
12	4 (12/13)	10	6.690.862	6.567.543	98,16%

*Figura 12: Resumen acertantes con más de 10 aciertos.*

Así mismo, como se puede comprobar en la página web, basándonos únicamente en las predicciones de nuestros usuarios, logramos pronosticar con una eficiencia superior a la mayoría como puede verse en la siguiente captura de pantalla representada en la *Figura 13*.

Posición	Usuario	Jornadas	Aciertos acumulados	Media de aciertos	Participación (%)
1		1	10	10	12.5
2		3	26	8.7	37.5
3		3	25	8.3	37.5
4	proyectoPetaso	8	64	8	100
5		2	16	8	25
6		1	8	8	12.5
7		1	8	8	12.5
8		5	38	7.6	62.5
9		7	53	7.6	87.5
10		2	15	7.5	25
11		4	30	7.5	50
12		3	22	7.3	37.5
13		5	36	7.2	62.5
14		8	57	7.1	100
15		2	14	7	25
16		6	42	7	75
17		4	28	7	50

*Figura 13: Ranking de las 8 primeras jornadas de la temporada 12/13.*

Como se puede observar nuestro algoritmo representado por el usuario “proyectoPetaso” se sitúa en 4º lugar por media de aciertos de los 36 usuarios que han participado con nosotros desde el comienzo de esta temporada 12/13. Cabe destacar que ostenta dicha posición, habiendo pronosticado las ocho jornadas posibles, mientras que los usuarios que o bien la superan o están empatados apenas han participado en una, dos o tres jornadas, teniendo en cuenta que lo que lastra a nuestra predicción son las dos primeras jornadas donde como ya hemos explicado anteriormente fueron de muy difícil predicción por sus resultados sorprendidos. Si no contáramos dichas jornadas, es decir, teniendo en cuenta las 6 últimas, la media de nuestra predicción ascendería hasta 9’33 aciertos por jornada, lo cual es un dato altamente positivo.

Otro dato a destacar es que la media global de aciertos por parte de los usuarios en la temporada actual incluyendo todas las jornadas, es de 6'685 aciertos por jornada, por lo que logramos obtener de media 1'315 aciertos más que nuestros usuarios como conjunto. Con lo que quedaría demostrado que el sistema funciona.

### 5.3.2 Datos de marketing y visitas web

Para poder llevar un control de las visitas y demás estadísticas de la web, usamos la herramienta Google Analytics, que nos ofrece amplios informes sobre los parámetros que le indiquemos. En el Anexo B podremos encontrar varios informes sobre estadísticas con numerosos datos de los visitantes web de nuestra página, como tiempo de navegación, páginas visitadas, dispositivo de acceso, país y ciudad de conexión, etc.

El gráfico que mostraremos a continuación en la *Figura 14* muestra las visitas diarias de nuestra web desde su lanzamiento el 9 de abril de 2012.



*Figura 14: Visitantes diarios de la web durante el periodo (09/04/12 – 14/10/12).*

Como se puede comprobar hubo un periodo de inactividad desde finales de junio a finales de agosto, correspondiente al cierre de apuestas de La Quiniela por falta de las competiciones futbolísticas necesarias. Si analizamos más en detenimiento el primer sector correspondiente al lanzamiento de la web hasta la finalización de la liga de primera división española el 13 de mayo de 2012. Podemos detectar un gran máximo que fue motivado por la difusión de la noticia sobre el lanzamiento de la web a través de diferentes redes sociales sobre todo a contactos personales de los componentes de

este proyecto. Si obviamos este gran máximo de 155 usuarios únicos el viernes 13 de abril de 2012, podemos observar en la siguiente gráfica de la *Figura 15*, hubo un flujo de visitantes semanales descendente debido muy posiblemente al olvido y/o falta de interés en la página.



*Figura 15: Visitantes diarios de la web durante el periodo (14/04/12 – 15/05/12).*

Por tanto, para evitar precisamente dicha pérdida de usuarios potenciales por culpa de los despistes personales, decidimos recurrir al envío de un correo electrónico corporativo de parte de la web donde anunciamos cada apertura de jornada, comentamos los resultados obtenidos en la jornada previa y animamos a la participación. Dicha estrategia la llevamos a cabo con el reinicio de la temporada 2012/2013. También cabe destacar que hemos detectado varios problemas a la hora de la recepción de estos emails por parte de los usuarios, ya que determinados servidores de correo lo clasifican como spam y por tanto suelen enviarlo a la bandeja de correo no deseado, por lo que no ha tenido toda la difusión que se pretendía y por tanto habrá minimizado los datos positivos.

La siguiente gráfica representada en la *Figura 16* que ilustra el periodo desde la semana previa al inicio de la liga de primera división española (19 de agosto de 2012), hasta el 14 de octubre de 2012. Como podemos observar, encontramos numerosos máximos destacados que corresponden al envío de los emails antes mencionados, por lo que creemos que es una herramienta de marketing útil y necesaria.





*Figura 16: Visitantes diarios de la web durante el periodo (12/08/12 – 14/10/12).*

Por otro lado cabe destacar que la semana previa a la jornada del 30 de septiembre de 2012, decidimos no difundir dicho correo para comprobar de una manera más o menos objetiva los resultados que obteníamos con y sin él, tanto de visitas a la web, como de participación en el pronóstico.

Los resultados no pudieron ser más claros, tuvimos un mínimo histórico de participación con apenas 8 usuarios y también el dato más bajo de visitas semanales desde nuestro comienzo. No obstante es reseñable que como se puede observar en la anterior gráfica, el martes 25 de septiembre de 2012, tuvimos un máximo local, lo que indica que varios usuarios habían entrado en una dinámica de juego en nuestra web, ya que se trataba del día en el que habitualmente procesamos la jornada y publicamos la nueva.

## 5.4 Corrección de sesgos en la recogida de información

En estadística, se llama *sesgo de un estimador* a la diferencia entre su esperanza matemática y el valor numérico del parámetro que estima. El no tener sesgo es una propiedad deseable de los estimadores. En nuestro caso, el sesgo es la diferencia entre el valor ideal del estimador, y el valor al que tiende debido a la mala interpretación de los pronósticos.

Inicialmente nuestro pronóstico se basaba únicamente en las nueve posibilidades repartidas entre los tres pronósticos 1, X y 2 dependiendo de su grado de confianza, dichos datos eran ponderados en media para obtener un resultado. Debido a esto nos encontramos con un gran problema, pues cuando un usuario no estaba para nada seguro de su pronóstico, tendía a marcar uno de las posibilidades de la X, ya que era la opción más imparcial, pero sin embargo sesgaba enormemente nuestro resultado hacia las X, y por tanto hacía inútil prácticamente el resto de los datos.

Fue por eso por lo que decidimos incluir la opción 0, que representa la opción de no saber con certeza indicar ninguno de los otros resultados. Por nuestra parte la opción del 0 la contabilizábamos como nula y no contaba para la media ponderada del algoritmo inicial, por lo que solucionamos dicho sesgo, o al menos esa era la intención, pues como hemos podido detectar y comprobar al contactar con algunos de ellos, la opción 0 es muy poco usada por el afán de intentar aportar algo al resultado, por lo que los usuarios por norma general aunque su razonamiento sea exactamente el de no saber qué resultado decir, o bien ponen un resultado al azar o bien se decantan de nuevo por la X. Por tanto creemos que dicho sesgo no se ha podido corregir de manera efectiva en la totalidad de los usuarios.

Una vez “corregido” dicho sesgo, continuamos observando una clara tendencia de los resultados hacia la X, y nos dimos cuenta de un gran error que solucionamos asignando cada una de las posibles opciones a probabilidades repartidas entre los tres resultados globales 1, X y 2 dadas por los diferentes grados de confianza como se observa en la tabla de la *Figura 7* mostrada anteriormente.

Dicho error se basaba en que al contraponer dos opiniones diferentes, estas no se contrarrestaban, si no que se sumaban y daban como resultado una tendencia hacia la X muy fuerte. Pongamos un par de ejemplos para ilustrarlo mejor:

- Si dos usuarios A y B, optan por las opciones extremistas de pronosticar un 1 y un 9 en un mismo partido, esto significaba que al hacer la media ponderada obteníamos un  $P_T = (P_A + P_B) / 2 = (1 + 9) / 2 = 5$  de media con dos usuarios, es decir, que era equivalente a que esos dos usuarios hubieran pronosticado un 5, es decir un empate. Por otro lado, con la nueva forma de cuantificar los pronósticos, el resultado de los usuarios habría sido el siguiente  $P_T[P_1, P_X, P_2] = (P_A[100, 0, 0] + P_B[0, 0, 100]) / 2 = P_T[50, 0, 50]$ , es decir, da un 50% de posibilidades al 1, y otro 50% al 2, dejando a la X sin probabilidad alguna, como es en el caso ideal, pues ninguno de los dos usuarios cree que pueda ocurrir dicho suceso y no tienden de manera equivocada fuertemente hacia la X.

- Si por otro lado, esos dos mismos usuarios A y B, optan por las opciones de pronosticar un 3 y un 7 en un mismo partido, esto significaba que al hacer la media ponderada obteníamos un  $P_T = (P_A + P_B) / 2 = (3 + 7) / 2 = 5$  de media con dos usuarios, es decir, que era equivalente a que esos dos usuarios hubieran pronosticado un 5, es decir un empate. Por otro lado, con la nueva forma de cuantificar los pronósticos, el resultado de los usuarios habría sido el siguiente  $P_T[P_1, P_X, P_2] = (P_A[55, 35, 10] + P_B[10, 30, 55]) / 2 = P_T[32.5, 35, 32.5]$ , es decir, prácticamente se anula el uno con el otro como era de esperar, pues cada uno piensa que ganará el equipo contrario al otro, pero dando muchas opciones al empate y alguna al contrincante, y al tener los tres pronósticos prácticamente la misma posibilidad es como si se anularan y no tienden de manera equivocada fuertemente hacia la X.

De esta forma, dando los pesos correspondientes apoyándonos en la *teoría Bayesiana*, conseguimos corregir dicho sesgo de manera eficiente debido a que la asignación de probabilidades de los tres sucesos 1, X y 2 es homogénea, no como de la manera inicial que el pronóstico de empate estaba sobre-representado.

Por último, un sesgo que se produce debido a la mala interpretación o utilización por parte de los usuarios de la interfaz, es el que se refiere al sesgo individual que efectúan algunos usuarios al basar sus resultados en pronósticos únicamente extremistas, es decir, únicamente pronostican, 1, 5 y 9, dependiendo si su opinión es que el resultado será un 1, una X o un 2. Dicho sesgo se produce o bien porque el usuario no comprende el mecanismo de decisión de pronósticos explicado junto al boleto de predicción, o bien porque está acérrimamente convencido de que sus pronósticos se van a cumplir al 100%. Por tanto una manera de intentar corregirlo sería mostrando de alguna manera los porcentajes equivalentes a cada una de las posibles opciones, cosa que no creemos conveniente como hemos explicado anteriormente por intentar simplificar lo máximo posible la decisión del usuario final.

## 5.5 Valoración de los mecanismos de los algoritmos de decisión colectiva a la vista de los datos

A la hora de valorar, qué aspectos hemos logrado o no cumplir en mayor o menor medida sobre los diferentes criterios que debe aunar un estadístico de inteligencia colectiva, haremos una autocrítica intentando ser lo más objetivos posibles evaluando los diferentes criterios que ha de tener dicho estadístico para lograr un correcto funcionamiento. Como ya expusimos y explicamos en el capítulo 2, los objetivos para lograr un buen estadístico de inteligencia colectiva son los siguientes:

- ***Mecanismo de agregación.*** Básicamente define la necesidad de tener un agregador o controlador, en este caso nosotros, que lleve un control de las predicciones obtenidas y combine esa información de manera que sea útil. Consideramos que este objetivo logramos cumplirlo, pues actuamos como un agregador individual e interpretamos los datos de una manera cuantificada que como hemos demostrado anteriormente es útil.
- ***Decisiones independientes.*** Argumenta que las decisiones de cada usuario han de ser independientes e incorrelacionadas con el respecto al resto de los usuarios, pues esto podría influir negativamente en la predicción de dicho usuario motivado por la presión de las demás decisiones del grupo. Logramos cumplir dicho objetivo gracias a que los participantes de nuestra predicción son auto-seleccionados, es decir, que ellos mismos son los que deciden participar; a su vez ningún usuario puede ver el resultado de los demás, de manera que no es influenciado por sus resultados.
- ***Multitud razonablemente “inteligente”.*** Este principio se basa en que el colectivo en general ha de tener los suficientes conocimientos sobre el problema que se plantea de manera que las opiniones de sus miembros no se basen únicamente en una cuestión de azar o sensaciones, si no que sepan interpretar las distintas variables concernientes al problema para llegar a una con-

clusión razonada sobre él. En nuestro caso en particular el colectivo debería de tener un cierto grado de conocimiento sobre todos los equipos participantes en los pronósticos, y siendo realistas, no logramos cumplir dicho objetivo pues muchos de ellos desconocen cualquier dato de la mayoría de los equipos más allá de su nombre y la ciudad a la que pertenece.

- ***Diversidad.*** Para lograr un buen estadístico es necesario reunir a un colectivo diverso, donde las opiniones de los individuos se contraponga para lograr una corrección de sesgos inducidos por una subjetividad parecida y abundante. Nuestro colectivo no logra este objetivo ya que la mayoría de sus usuarios están centralizados geográficamente en Madrid (datos obtenidos con *Google Analytics* incluidos en el Anexo B), lo que creemos es un hándicap ya que la mayoría de ellos están vinculados emocionalmente con los equipos de la ciudad.
- ***Complejidad.*** Anécdotas de inteligencia colectiva sugieren que esta obtiene mejores resultados cuanto más complejo es un problema, ya que ayuda a evaluar todas las cuestiones que lo conciernen de manera más amplia que un individuo concreto. En nuestro caso, el de la predicción de resultados de partidos de fútbol, podría aplicarse dicha teoría, ya que las variables que influyen en cada resultado son numerosas, como por ejemplo el presupuesto de los equipos, clasificación actual, racha de resultados recientes, estado físico de sus jugadores clave, etc.
- ***Descentralización.*** Es la propiedad de un sistema en el que las decisiones son tomadas por individuos basándose en su propio conocimiento local y específico en vez de por un planificador omnisciente. En nuestro caso también cumplimos dicha propiedad, ya que cada usuario hace su predicción en base a su conocimiento individual y la toma de decisiones es meramente cuantitativa respecto a los pronósticos del conjunto de usuarios.

Por tanto podemos concluir que en su amplia mayoría cumplimos los criterios para lograr un buen estadístico de inteligencia colectiva, pero cabe destacar, que si lográramos obtener tanto una multitud razonablemente “inteligente”, futbolísticamente hablando, como diversa en aspectos geográficos y de simpatización de clubs de fútbol, lograríamos probablemente una mejora en la toma de decisiones.

# CAPÍTULO 6

## CONCLUSIONES Y DISCUSIÓN

El experimento realizado en este proyecto nace con la intención de valorar la eficacia de la decisión colectiva aplicada al caso particular de las quinielas de fútbol. Se demuestra, que algoritmos simples de agregación, son capaces de generar predicciones con más precisión de las que las personas individuales realizan, siendo estas la entrada a nuestro sistema de agregación. Para potenciar dicha precisión, se combina la decisión colectiva con estadísticos bayesianos determinados por una matriz de probabilidades que se aplican para dar un peso conveniente a cada estimación, de manera que los pensamientos de cada persona sobre un resultado se vean reflejados de manera cuantitativa en probabilidades a cada suceso.

Dicho experimento también sugiere que la interfaz puede sesgar la recogida de datos como se expuso en el capítulo 5, ya que, por una parte sigue estando incompleta la cuantización uniforme sobre el total del área de probabilidades, y por otra debido a la mala interpretación de los usuarios al pronosticar sus predicciones siempre de manera extremista. Por tanto creemos fundamental hacer un uso práctico de dicha interfaz, intentando asemejar el pensamiento real sobre un pronóstico a las indicaciones sugeridas en la página, las cuales creemos reflejan más fielmente la matriz de probabilidades usada.

Aunque es difícil valorar de forma objetiva la eficacia de la predicción, la comparación con sistemas alternativos basados en procesos estructurados puede servirnos de guía. Como una conclusión comparativa, podemos destacar el mejor resultado obtenido con respecto al estudio de Díaz Molk en 2002 en la Universidad de Guadalajara (México) descrito con anterioridad en el capítulo 1 (3). Dicho estudio basaba su decisor en un perceptrón multicapa aplicado a las RNA apoyándose en datos históricos tales como: presupuesto de los equipos, clasificación actual, puntos conseguidos en los últimos cuatro partidos, goles anotados y encajados en los últimos cuatro partidos, etc. El resultado de dicho estudio demostró que pronosticando resultados de los partidos de fútbol únicamente de primera división de la LFP durante cuatro jornadas consecutivas, acertaban de media el 53'75% de los resultados. Si marginamos nuestros resultados obtenidos únicamente con los partidos de primera división durante la presente temporada 2012/2013, obtenemos una tasa de acierto del 60'3% al pronosticar con éxito 38 de los 63 resultados posibles.

De manera añadida, si consideramos que se seleccionaron las cuatro jornadas consecutivas donde los resultados fueron mejores de cara al estudio y aplicamos el mismo criterio, obtendríamos una tasa de acierto de 72'22% al pronosticar con éxito 26 de los 36 resultados posibles evaluando las jornadas 3, 4, 5 y 6 de la presente temporada.

En una segunda valoración entre sistemas de predicción colectiva, comparamos el uso del estadístico bayesiano con la media ponderada de las estimaciones independientes de los usuarios. Analizando los resultados de las primeras ocho jornadas de la presente temporada con ambos métodos, llegamos a la conclusión de que el estadístico bayesiano logra una predicción mucho más efectiva debido a la eliminación de sesgos que conlleva su uso, ya que obtenemos una media de 8 aciertos por jornada, mientras que con la media ponderada de las estimaciones de los usuarios este dato es de 6'5 aciertos por jornada.

Cabe destacar de manera adicional, que la media de los aciertos obtenidos por todos los usuarios durante las mencionadas ocho primeras jornadas es de 6'493 aciertos por jornada, prácticamente idéntica al resultado de los aciertos del sistema que basa su decisión en la media ponderada de las estimaciones de los usuarios.



Desde el punto de vista del marketing y utilizando datos objetivos como son los visitantes a nuestra página web así como los participantes de forma activa en nuestra predicción, podemos extraer como principal conclusión que las redes sociales son un elemento muy importante a la hora de dar a conocer un nuevo servicio, pues en pocas horas tras difundir la apertura de nuestra web, pudimos experimentar un aumento exponencial en el número de visitas y por tanto en el registro de usuarios así como de la participación en nuestra interfaz de predicciones.

Cabe destacar que tras el incremento inicial, tanto las visitas a la web como los usuarios que participaban en cada jornada fue disminuyendo paulatinamente hasta el cese de actividad de la página debido al parón veraniego de apuestas de La Quiniela. Dicha disminución se produjo probablemente por los malos resultados iniciales así como de manera especial por el olvido de los usuarios o la falta de interés.

Para poder paliar que dichos usuarios no recordaran introducir sus pronósticos aun teniendo la intención de hacerlo, decidimos mandar un correo electrónico en la semana previa a cada jornada con buenos resultados, ya que como pudimos comprobar, se incrementaron notablemente tanto las visitas como las predicciones realizadas tras cada uno de estos recordatorios.

En un futuro, sería recomendable interactuar más con el usuario a través de las redes sociales y otros medios gratuitos de difusión, así como crear una interfaz aún más novedosa e impactante a ojos del usuario. No obstante, no cabe duda de que unos muy buenos resultados con los que la web obtuviera beneficios por sus predicciones ayudaría de manera más que notable a una participación masiva, evento que creemos bastante improbable.

Por tanto concluimos que en el caso propuesto de la predicción de resultados en partidos de primera división de la LFP, obtiene mejores resultados un algoritmo de decisión colectiva basado en estadísticos bayesianos que un sistema de decisión de redes neuronales basado en un perceptrón multicapa utilizando datos históricos.

Así mismo, se puede concluir que nuestros diferentes criterios basados decisión colectiva son capaces de predecir con mayor precisión un suceso de manera notable, en nuestro caso pronósticos de La Quiniela, que un conjunto de usuarios tanto de manera individual como por decisión colectiva a través de la media ponderada de sus predicciones.

# APÉNDICE A

## ESQUEMA DE LA WEB

El portal web implementado ha sido desarrollado íntegra y exclusivamente para este proyecto sin ningún tipo de plantilla ni método prediseñado, por lo que su diseño y funcionalidades son producto del desarrollo del mismo.

Dicho portal se divide en secciones funcionales, diferenciándolas de la siguiente manera:

- **Registro/autenticación de usuario.** Esta sección situada en la cabecera de la web, posibilita el registro de nuevos usuarios en la BDD, así como permite el acceso a las secciones privadas en caso de mostrar los credenciales correctos. De manera adicional existe un servicio de recordatorio de usuario y contraseña a través de correo electrónico para aquellos usuarios que no los recuerden.
- **Navegación por la web.** El menú lateral localizado a la izquierda de la página, posibilita el acceso a los diferentes contenidos de la web, tales como información, ranking de los usuarios e incluso tareas de administración. Dicho menú se comporta de forma dinámica dando acceso a los diferentes contenidos al usuario que lo visualiza dependiendo de los permisos que este posea.

Un usuario no registrado, apenas tendrá acceso a la información de la web y resultados anteriores obtenidos por nuestro sistema, mientras que adicionalmente, los usuarios registrados tendrán acceso tanto a sus datos personales como al ranking general y de cada jornada con estadísticas sobre todos los usuarios. Aquellos usuarios que posean los permisos de administrador, también podrán acceder a los resultados elaborados por los diferentes sistemas de decisión implementados en todo momento, así como a secciones de administración del sistema.

- **Sección principal de la web.** Es la parte dinámica de dicha web donde se muestran los diferentes contenidos en función de la página que se está visitando. Así pues en dicha sección se muestra tanto información, como tablas con resultados o el interfaz de recogida de pronósticos. Dichas informaciones e interfaces, varían dependiendo del usuario que las visualice, ya que acceden a información personal; de la misma manera, dependiendo en qué momento de la jornada se consulte, se mostrará el interfaz de recogida de datos o los resultados del pronóstico de la jornada en juego.
- **Contacto.** Por último cabe destacar la posibilidad de los usuarios de enviar cualquier tipo de información a los administradores de la web a través del enlace situado en el pie de esta.

A continuación mostramos en la *Figura 17*, el esquema de la web donde se describe la interacción entre las diferentes secciones de esta:

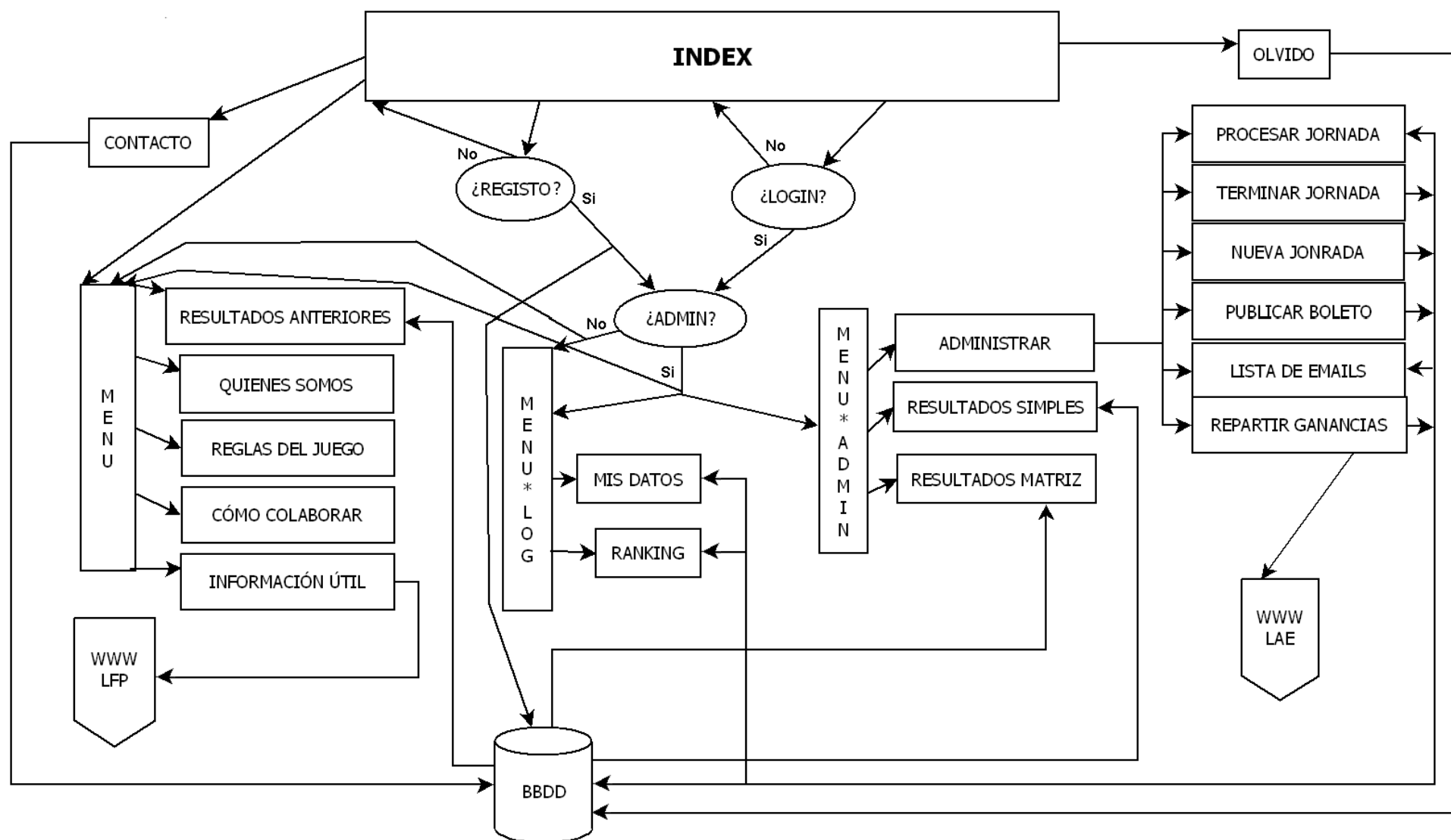


Figura 17: Esquema de *www.petaso1x2.com*.

A continuación describiremos de manera sencilla cada página y sección de la web mostrados en el esquema:

- **Index:** página principal del portal web donde se ubica el interfaz de recogida de pronósticos. Está accesible desde cualquier página de la web a través de la cabecera.
- **Registro:** página donde se genera el registro de usuarios en el sistema.
- **Login:** sección de la cabecera donde un usuario registrado puede autenticarse y entrar como tal en el sistema. Adicionalmente se comprueba si dicho usuario tiene privilegios de administrador para habilitar las secciones destinadas a ello.
- **Olvido:** página donde un usuario puede comunicar que desea recordar su contraseña y/o usuario a través de su correo electrónico de registro.
- **Menú (básico):** sección donde se encuentran ubicados los diferentes accesos a las páginas de la web que tiene permiso el usuario existente. El menú sin privilegios da acceso a las páginas: *Resultados anteriores*, *Quienes somos*, *Reglas del juego*, *Cómo colaborar* e *Información útil*.
- **Menú log:** opciones de navegación adicionales que se suman al menú básico para usuarios autenticados en el sistema. Dichas opciones son: *Mis datos* y *Ranking*.
- **Menú admin:** opciones de navegación adicionales al Menú log para los usuarios con privilegios de administrador. Dichas opciones son: *Administrar*, *Resultados simples* y *Resultados Matriz*.
- **Resultados anteriores:** página en la que se muestra la comparativa de el resultado pronosticado por nuestro sistema con los resultados reales de cada jornada. Adicionalmente si accede un usuario autenticado, se mostrarán también sus resultados para cada una de dichas jornadas.
- **Quienes somos:** página en la que se define el propósito del portal web.

- **Reglas del juego:** página en la que se exponen los criterios en los que se basa la participación en la interfaz.
- **Cómo colaborar:** página en la que se muestran las diferentes maneras de colaborar en el pronóstico de predicciones.
- **Información útil:** Página en la que se muestran enlaces web externos con información útil para el usuario, cómo es la página web de la *LFP*.
- **Mis datos:** página en la que se muestran los datos personales del usuario.
- **Ranking:** página dinámica en la que se muestran tablas de clasificación por aciertos de cada jornada jugada, así como una clasificación general de todos los usuarios con diferentes estadísticas.
- **Administrar:** página que da acceso a las diferentes acciones de administración de la web, como son: *Procesar jornada*, *Terminar jornada*, *Nueva jornada*, *Publicar boleto*, *Lista de emails* y *Repartir ganancias*.
- **BBDD:** Bases de datos de almacenamiento de información de la web.
- **Procesar jornada:** página en la que una vez finalizada la jornada se introducen los resultados finales y procesa los datos de manera automática actualizando los aciertos obtenidos por cada usuario.
- **Terminar jornada:** Deshabilita la posibilidad de introducir una predicción en la interfaz de recogida de datos. Se ejecuta unas horas antes del comienzo de la jornada para poder elaborar el pronóstico correspondiente.
- **Nueva jornada:** página en la que se introducen los equipos que van a formar parte de la nueva jornada y habilita la interfaz de introducción de pronósticos.
- **Publicar boleto:** página en la que una vez comenzados los partidos de la jornada, se introduce el pronóstico elaborado por el sistema y se publica en la web.

- **Lista de emails:** Muestra por pantalla la lista completa de los emails de los usuarios registrados.
- **Repartir ganancias:** página en la que se introduce la cantidad ganada en una jornada y la reparte en el saldo virtual de la BDD a cada usuario que haya participado. Así mismo contiene el link externo a la página oficial de *LAE* para la comprobación de dichas ganancias.
- **Contacto:** página en la que se ofrece al usuario la posibilidad de contactar con los administradores del sistema.



# APÉNDICE B

## INFORMES DE GOOGLE ANALYTICS

### B.1 Duración y número de páginas vistas por visita



Proyecto Petaso-<http://petaso1x2.com/>  
Proyecto Petaso PREDETERM...

09/04/2012 - 14/10/2012

#### Interacción

% de visitas: 100,00%

Rendimiento

Duración de la visita

Visitas		Páginas vistas	
<b>1.511</b>		<b>9.476</b>	
% del total: 100,00% (1.511)		% del total: 100,00% (9.476)	
Duración de la visita	Visitas	Páginas vistas	Porcentaje del total
			■ Visitas ■ Páginas vistas
0-10 segundos	448	536	29,85% 5,66%
11-30 segundos	116	327	7,68% 3,45%
31-60 segundos	120	522	7,94% 5,51%
61-180 segundos	310	1.732	20,52% 18,28%
181-600 segundos	301	2.433	19,92% 25,68%
601-1800 segundos	136	1.764	9,00% 18,62%
1801+ segundos	80	2.162	5,29% 22,82%

© 2012 Google

## B.2 Navegadores y SSOO



Proyecto Petaso-<http://petaso1x2.com/>  
Proyecto Petaso PREDETERM...

### Navegador y SO

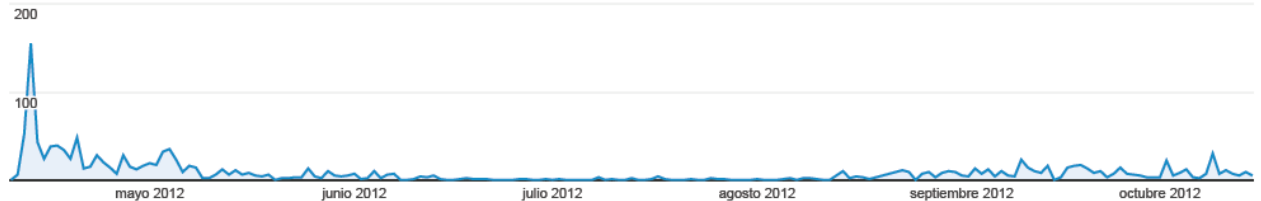
10/04/2012 - 15/10/2012

100% de visitas: 100,00%

Explorador

Uso del sitio

Visitas



Visitas

**1.517**

% del total: 100,00% (1.517)

Páginas / Visita

**6,26**

Promedio del sitio: 6,26 (0,00%)

Duración media de la visita

**00:06:17**

Promedio del sitio: 00:06:17 (0,00%)

Porcentaje de visitas nuevas

**34,94%**

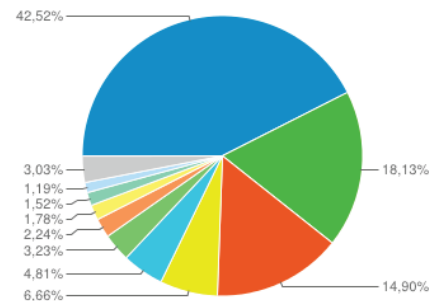
Promedio del sitio: 34,94% (0,00%)

Porcentaje de rebote

**25,38%**

Promedio del sitio: 25,38% (0,00%)

Navegador	Sistema operativo	Visitas	Visitas	Contribución al total: Visitas
1. Chrome	Windows	645	42,52%	
2. Internet Explorer	Windows	275	18,13%	
3. Firefox	Windows	226	14,90%	
4. Android Browser	Android	101	6,66%	
5. Safari	Macintosh	73	4,81%	
6. Safari	iOS	49	3,23%	
7. Safari	iPhone	34	2,24%	
8. Firefox	Linux	27	1,78%	
9. Chrome	Linux	23	1,52%	
10. Chrome	Macintosh	18	1,19%	
11. Mozilla Compatible Agent	iPhone	16	1,05%	
12. Safari	iPad	11	0,73%	
13. Firefox	Android	6	0,40%	
14. Firefox	Macintosh	5	0,33%	
15. Safari	BlackBerry	3	0,20%	
16. Mozilla Compatible Agent	iPad	2	0,13%	
17. IE with Chrome Frame	Windows	1	0,07%	
18. Safari	Linux	1	0,07%	
19. Safari	SymbianOS	1	0,07%	



## B.3 Dispositivos móviles



Proyecto Petaso-<http://petaso1x2.com/>  
Proyecto Petaso PREDETERM...

### Dispositivos

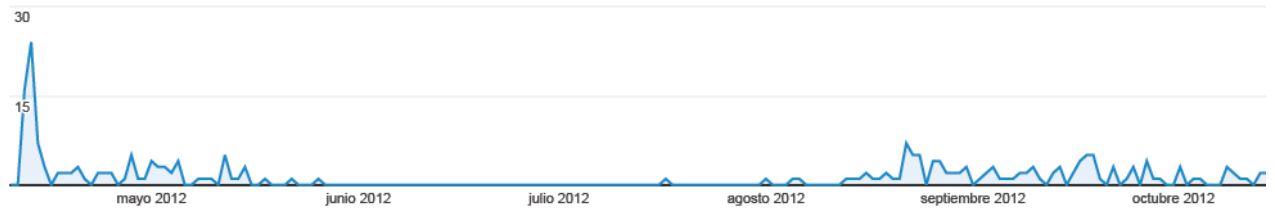
10/04/2012 - 15/10/2012

▼ % de visitas: 14,70%

Explorador

Uso del sitio

● Visitas



Visitas

**223**

% del total: 14,70% (1.517)

Páginas / Visita

**3,63**

Promedio del sitio: 6,26 (-42,04%)

Duración media de la visita

**00:02:14**

Promedio del sitio: 00:06:17 (-64,51%)

Porcentaje de visitas nuevas

**29,60%**

Promedio del sitio: 34,94% (-15,29%)

Porcentaje de rebote

**40,81%**

Promedio del sitio: 25,38% (60,79%)

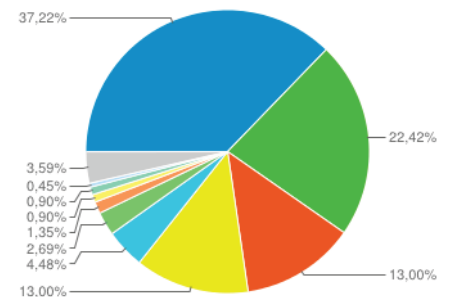
Información del dispositivo móvil

Visitas

Visitas

Contribución al total: Visitas

1.	Apple iPhone	83	37,22%
2.	Samsung GT-I9100 Galaxy S II	50	22,42%
3.	Apple iPad	29	13,00%
4.	SonyEricsson LT15i Xperia Arc	29	13,00%
5.	(not set)	10	4,48%
6.	Samsung GT-I9003	6	2,69%
7.	Samsung GT-I9000 Galaxy S	3	1,35%
8.	Samsung GT-S5830 Galaxy Ace	2	0,90%
9.	SonyEricsson MT15i Xperia Neo	2	0,90%
10.	Fujitsu T-01C REGZA Phone T-01C	1	0,45%



Filas 1 - 10 de 18

© 2012 Google

## B.4 Idiomas



Proyecto Petaso-<http://petaso1x2.com/>  
Proyecto Petaso PREDETERM...

### Información de visitantes

09/04/2012 - 14/10/2012

● % de visitas: 100,00%

#### Visión general

● Visitas



529 usuarios han visitado este sitio.

Visitas: 1.511

Visitantes exclusivos: 529

Páginas vistas: 9.476

Páginas / Visita: 6,27

Duración media de la visita: 00:06:02

Porcentaje de rebote: 25,28%

Porcentaje de visitas nuevas: 35,01%



■ 64,92% Returning Visitor

981 Visitas

■ 35,08% New Visitor

530 Visitas

Idioma		Visitas	% Visitas
1.	es	943	62,41%
2.	es-es	434	28,72%
3.	en	67	4,43%
4.	en-us	65	4,30%
5.	c	1	0,07%
6.	en_us	1	0,07%

[ver todo el informe](#)

© 2012 Google

## B.5 Países



Proyecto Petaso-http://petaso1x2.com/  
Proyecto Petaso PREDETERM...

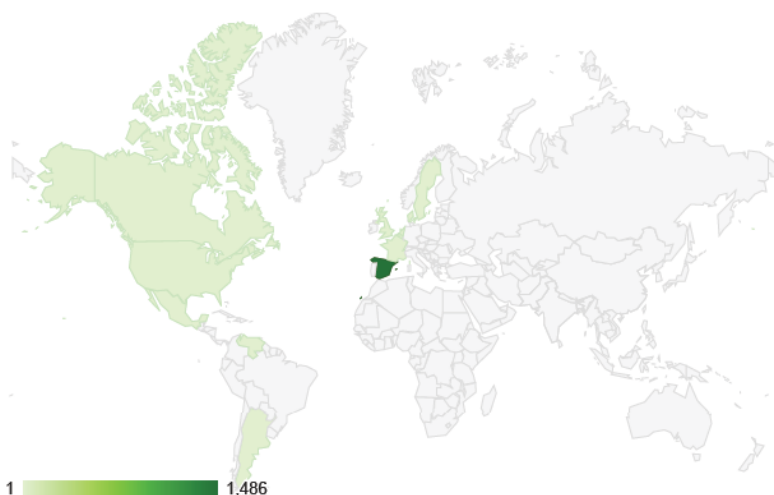
### Ubicación

09/04/2012 - 14/10/2012

% de visitas: 100,00%

Gráfico de visitas por ubicación

Uso del sitio



Visitas	Páginas / Visita	Duración media de la visita	Porcentaje de visitas nuevas	Porcentaje de rebote
<b>1.511</b>	<b>6,27</b>	<b>00:06:02</b>	<b>35,01%</b>	<b>25,28%</b>
% del total: 100,00% (1.511)	Promedio del sitio: 6,27 (0,00%)	Promedio del sitio: 00:06:02 (0,00%)	Promedio del sitio: 35,01% (0,00%)	Promedio del sitio: 25,28% (0,00%)

Pais/territorio	Visitas	Páginas / Visita	Duración media de la visita	Porcentaje de visitas nuevas	Porcentaje de rebote
1. <a href="#">Spain</a>	1.486	6,33	00:06:07	34,52%	24,90%
2. <a href="#">United States</a>	7	1,86	00:00:54	100,00%	57,14%
3. <a href="#">France</a>	3	5,67	00:01:55	33,33%	0,00%
4. <a href="#">United Kingdom</a>	3	4,33	00:01:13	100,00%	33,33%
5. <a href="#">Belgium</a>	2	5,50	00:02:57	0,00%	0,00%
6. <a href="#">Denmark</a>	2	1,00	00:00:00	50,00%	100,00%
7. <a href="#">Sweden</a>	2	2,50	00:07:18	0,00%	50,00%
8. <a href="#">Venezuela</a>	2	1,00	00:00:00	100,00%	100,00%
9. <a href="#">Argentina</a>	1	1,00	00:00:00	100,00%	100,00%
10. <a href="#">Canada</a>	1	1,00	00:00:00	100,00%	100,00%
11. <a href="#">Mexico</a>	1	2,00	00:00:26	100,00%	0,00%
12. <a href="#">Netherlands</a>	1	2,00	00:00:14	0,00%	0,00%

Filas 1 - 12 de 12

## B.6 Ciudades



Proyecto Petaso-<http://petaso1x2.com/>  
Proyecto Petaso PREDETERM...

10/04/2012 - 15/10/2012

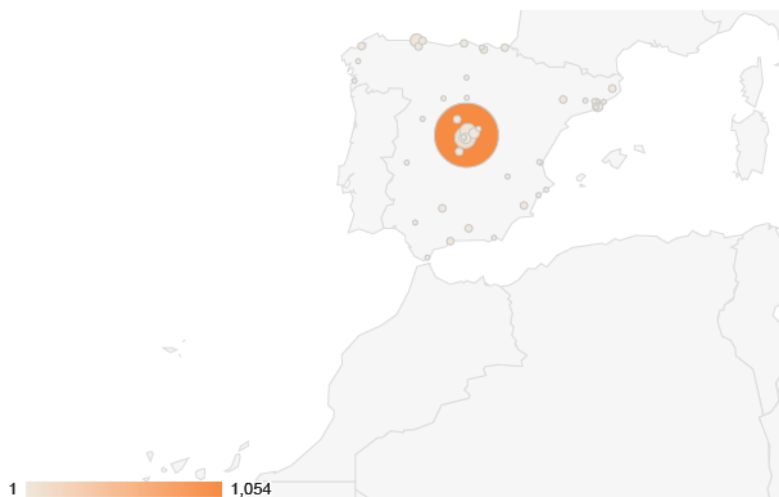
### Ubicación

TODAS » PAÍS/TERRITORIO: Spain

% de visitas: 98,35%

Gráfico de visitas por ubicación

Uso del sitio



Visitas	Páginas / Visita	Duración media de la visita	Porcentaje de visitas nuevas	Porcentaje de rebote
<b>1.492</b>	<b>6,32</b>	<b>00:06:22</b>	<b>34,45%</b>	<b>25,00%</b>
% del total: 98,35% (1.517)	Promedio del sitio: 6,26 (0,94%)	Promedio del sitio: 00:06:17 (1,28%)	Promedio del sitio: 34,94% (-1,39%)	Promedio del sitio: 25,38% (-1,49%)

Ciudad	Visitas	Visitas	Contribución al total:
1.  Madrid	1.054	70,64%	
2.  Leganes	110	7,37%	
3.  Alcobendas	48	3,22%	
4.  Aviles	34	2,28%	
5.  Barcelona	24	1,61%	
6.  Getafe	17	1,14%	
7.  Alcala de Henares	15	1,01%	
8.  Bilbao	14	0,94%	
9.  Lerida	14	0,94%	
10.  Segovia	13	0,87%	

Filas 1 - 10 de 48

# Bibliografía

1. Ministerio de Economía y Hacienda. BOE. [En línea] 23 de Julio de 2009. [http://www.loteriasypuestas.es/uploads/documentos/documentos\\_Rs6-7-09laquiniela\\_83cac456.pdf](http://www.loteriasypuestas.es/uploads/documentos/documentos_Rs6-7-09laquiniela_83cac456.pdf). 177.
2. Choque Aspiazu, Guillermo. eldiario.net. [En línea] Agosto de 2010. <http://www.eldiario.net/computacion/7-100816/index.htm>.
3. Díaz Molk, Marcos. Predicción de quinielas futbolísticas mediante un perceptrón multicapa. *Universidad de Guadalajara, Mexico*. [En línea] Febrero de 2002. <http://proton.ucting.udg.mx/posgrado/cursos/idc/aplicaciones/quinelas/index.html>.
4. Wikipedia. [En línea] 2012. [http://es.wikipedia.org/wiki/Toma\\_de\\_decisiones#Procesos\\_cognitivos\\_implicados\\_en\\_la\\_toma\\_de\\_decisiones](http://es.wikipedia.org/wiki/Toma_de_decisiones#Procesos_cognitivos_implicados_en_la_toma_de_decisiones).
5. Wikipedia. [En línea] 2012. [http://es.wikipedia.org/wiki/Inferencia\\_bayesiana](http://es.wikipedia.org/wiki/Inferencia_bayesiana).
6. M. Bishop, Christopher. Bayesian probabilities. *Pattern Recognition and Machine Learning*. Cambridge : Microsoft Research Ltd, 2006.
7. Cox, R. T. Probability, frequency and reasonable expectation. *American Journal of Physics*. 14, 1946, Vol. 1, 1-13.
8. Jaynes, E. T. Probability Theory: The Logic of Science. *Cambridge University Press*. 2003.
9. Watkins, Jennifer H. Prediction Markets as an Aggregation Mechanism for Collective Intelligence. *Human Complex Systems*. UC Los Angeles : s.n., 2007.

10. Heylighen, F. Collective intelligence and its implications of the web: algorithms. *Computational and Mathematical Organization*. 1999.
11. Surowiecki, J. *The wisdom of crowds*. New York : Doubleday, 2004.
12. Hogg, M. A., & Abrams, D. Intergroup behavior and social identity. *The SAGE handbook of social psychology*. 2003.
13. Worchel, S. Come one, come all: Toward understanding the process of collective. *The SAGE handbook of social psychology*. 2003.
14. Martin, R., & Hewstone, M. Social influence processes of control and change: Conformity,. *The SAGE handbook of social psychology*. 2003.
15. Watkins, J. H. *Individuals unite!: Generating synergistic understanding through online prediction markets*. 2005.
16. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. *Evaluating collaborative*. 2004.
17. Rodriguez, M. A. Social decision making with multi-relational networks and grammarbased. *Annual Hawaii International Conference on Systems*. Waikoloa, Hawaii : s.n., 2007.
18. Wade, C., & Tavis, C. *Invitation to psychology*. 2002.
19. Williamson, O. E. The economics of organization: The transaction cost approach. *American Journal of Sociology*. 1981, 87.
20. Hong, L., & Page, S. E. Problem solving by heterogeneous agents. [En línea] 1998. <http://ishi.lanl.gov/documents/hong.and.page1.pdf>.
21. librosweb.es. [En línea] <http://www.librosweb.es/xhtml/capitulo1.html>.
22. Definición abc. [En línea] <http://www.definicionabc.com/tecnologia/mysql.php>.
23. PHP.net. [En línea] <http://www.php.net/manual/es/intro-what-is.php>.
24. Wikipedia. [En línea] [http://es.wikipedia.org/wiki/An%C3%A1lisis\\_web](http://es.wikipedia.org/wiki/An%C3%A1lisis_web).
25. Wikipedia. [En línea] [http://es.wikipedia.org/wiki/Navegador\\_web](http://es.wikipedia.org/wiki/Navegador_web).